

SEGMENT ROUTING FOR SDN

Shaowen Ma, APAC Product Director, Juniper, mashao@juniper.net

March 1, 2017

AGENDA

Introduction

Segment Routing Deep Dive

Segment Routing SDN and Use Case

Summary

MPLS – 16 YEARS, GREAT SUCCESS

THE ACTUAL STANDARD FOR SERVICE DELIVERY

- LDP, mLDP
- RSVP-TE, RSVP-TE P2MP
- L3 MPLS VPN
- 6VPE/6PE
- L2 MPLS VPN – VPWS
- L2 MPLS VPN – VPLS (LDP, BGP, BGP AD)
- Next-generation multicast VPN
- MPLS-OAM, LSP BFD, VCCV Ping, and VCCV-BFD
- MPLS-TP Static LSP/PW, OAM, APS
- GMPLS, GMPLS UNI*



Kireeti Kompella



Eric Rosen



Yakov Rekhter

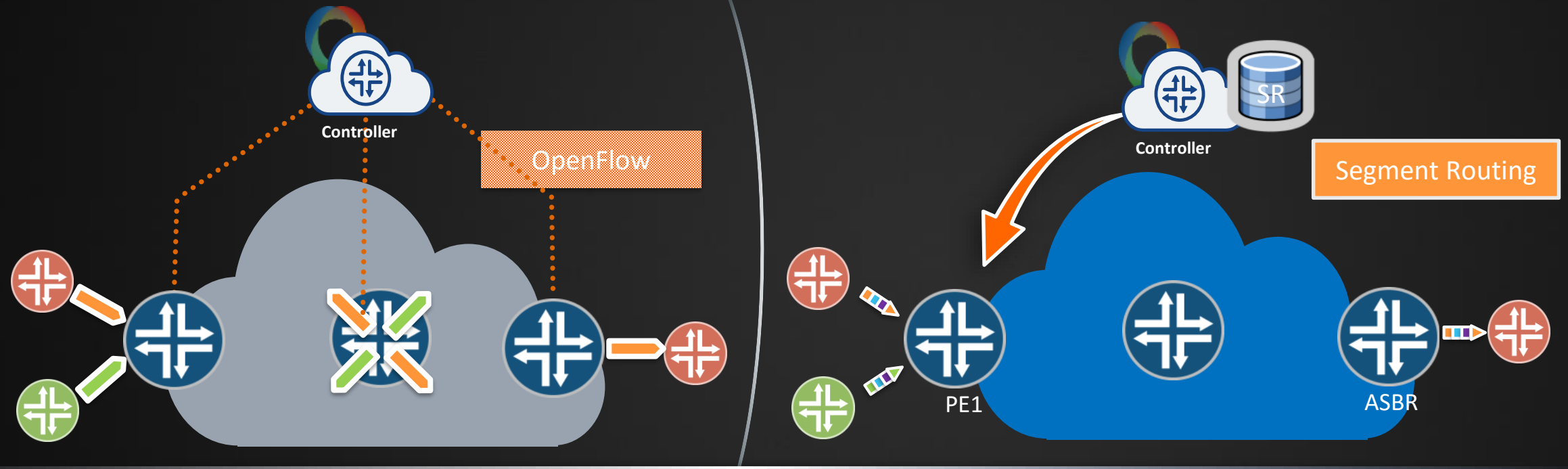
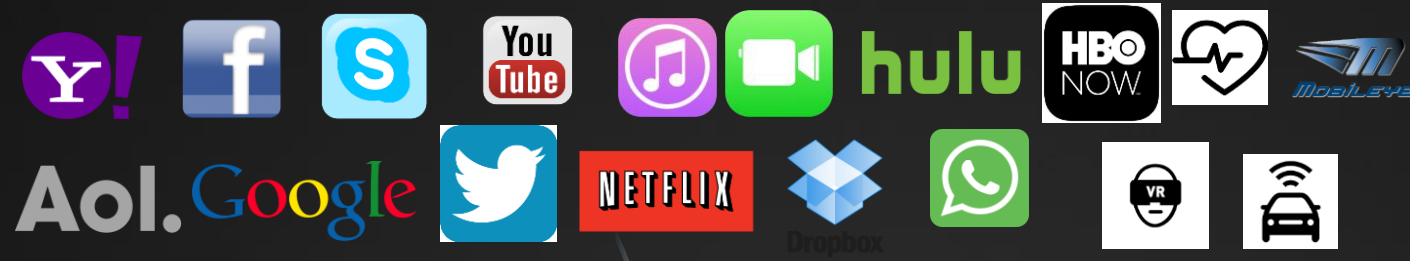


Many...

IETF SPRING/Segment Routing working group

- Source Packet Routing in Networking

SDN 2.0 ERA



Segment Routing, RSVP-TE Enable SDN 2.0
Edge Intelligence, Stateless CORE

AGENDA

Introduction

Segment Routing Deep Dive

Segment Routing SDN and Use Case

Summary

Segment Routing Introduction

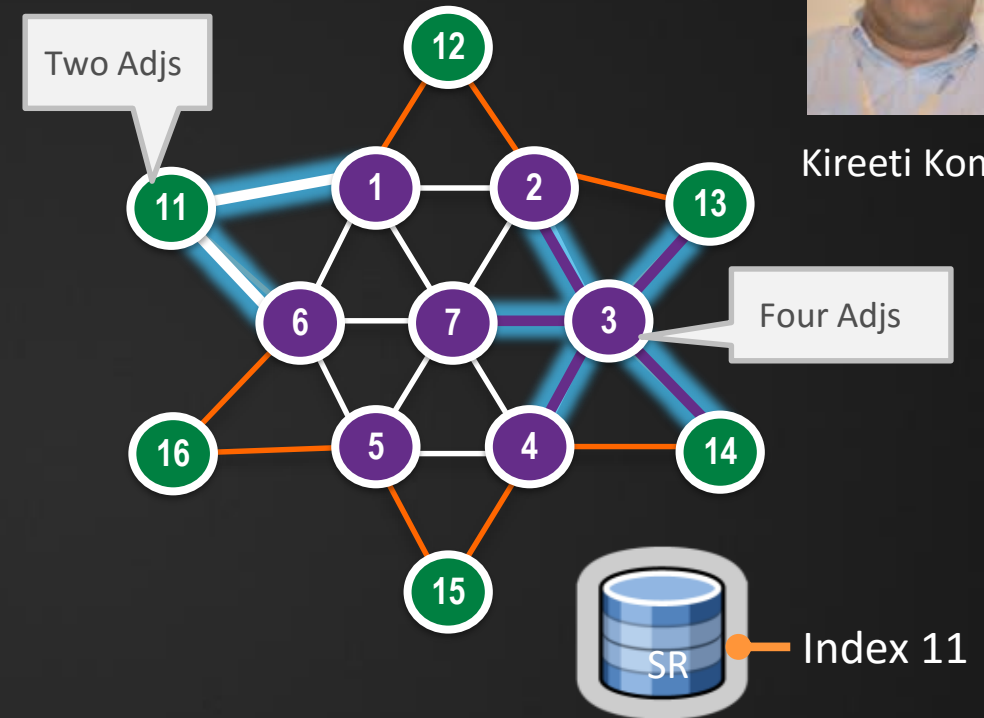
Source Based Routing

draft-ietf-isis-segment-routing-extensions-xx



Kireeti Kompella

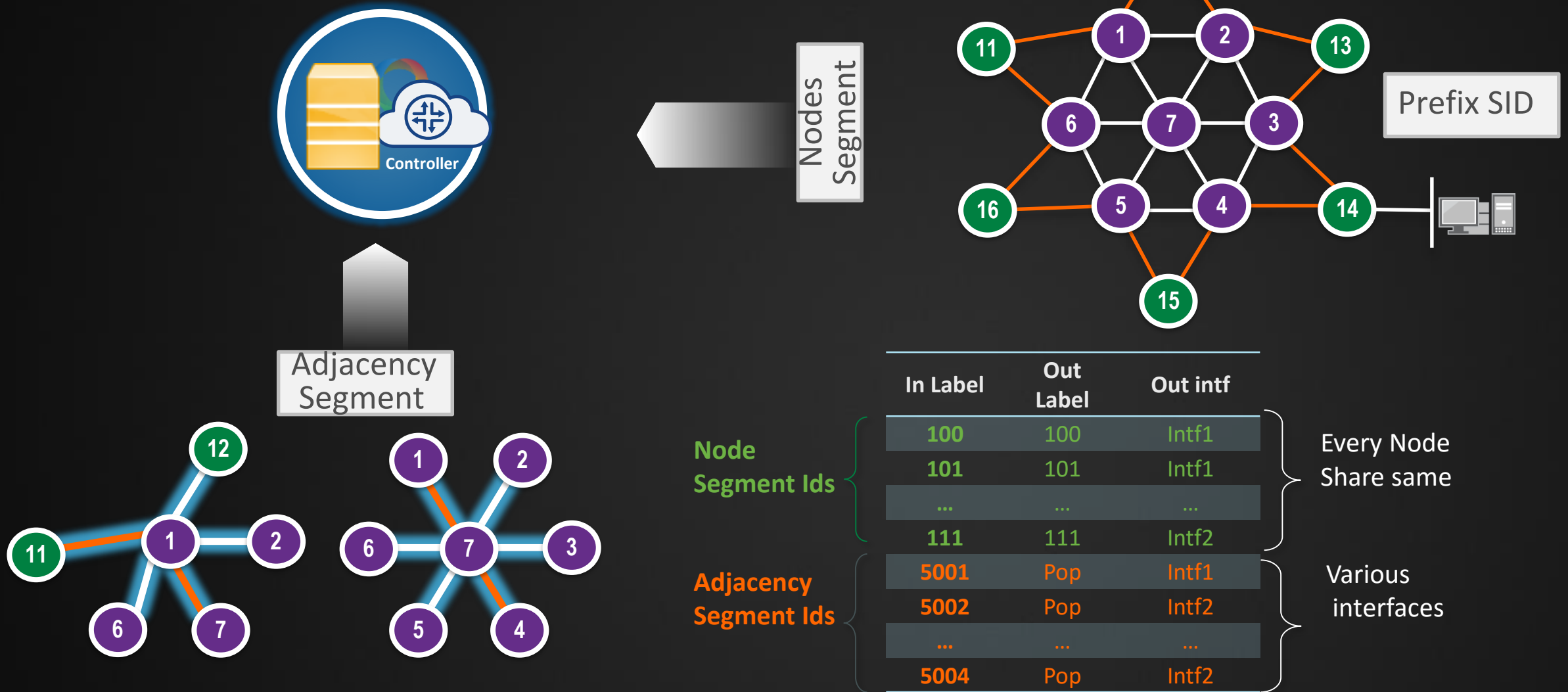
- Idea from Draft-Kompella(Label Block and Index)
- Network represented by Segment
 - Adj, Nodal Segment(unique #, one segment)
 - Segments act as topological sub-paths that can be combined together to form the desired path.
 - Source Routing: the source chooses a path and encodes it in the packet header as an ordered list of segments
- Every Node Forwarding table only take care portion of network
 - All nodal segment, SRGB(SR Global Block)
 - Adj Segment, No neighbors Adj Segment, Local Significant
- CSPF for nodal Segment
 - Calculate the OIF only,
 - label keep same(64-5000 reserved)



```
protocols { isis {  
  source-packet-routing { node-segment ipv4-index 11}}
```

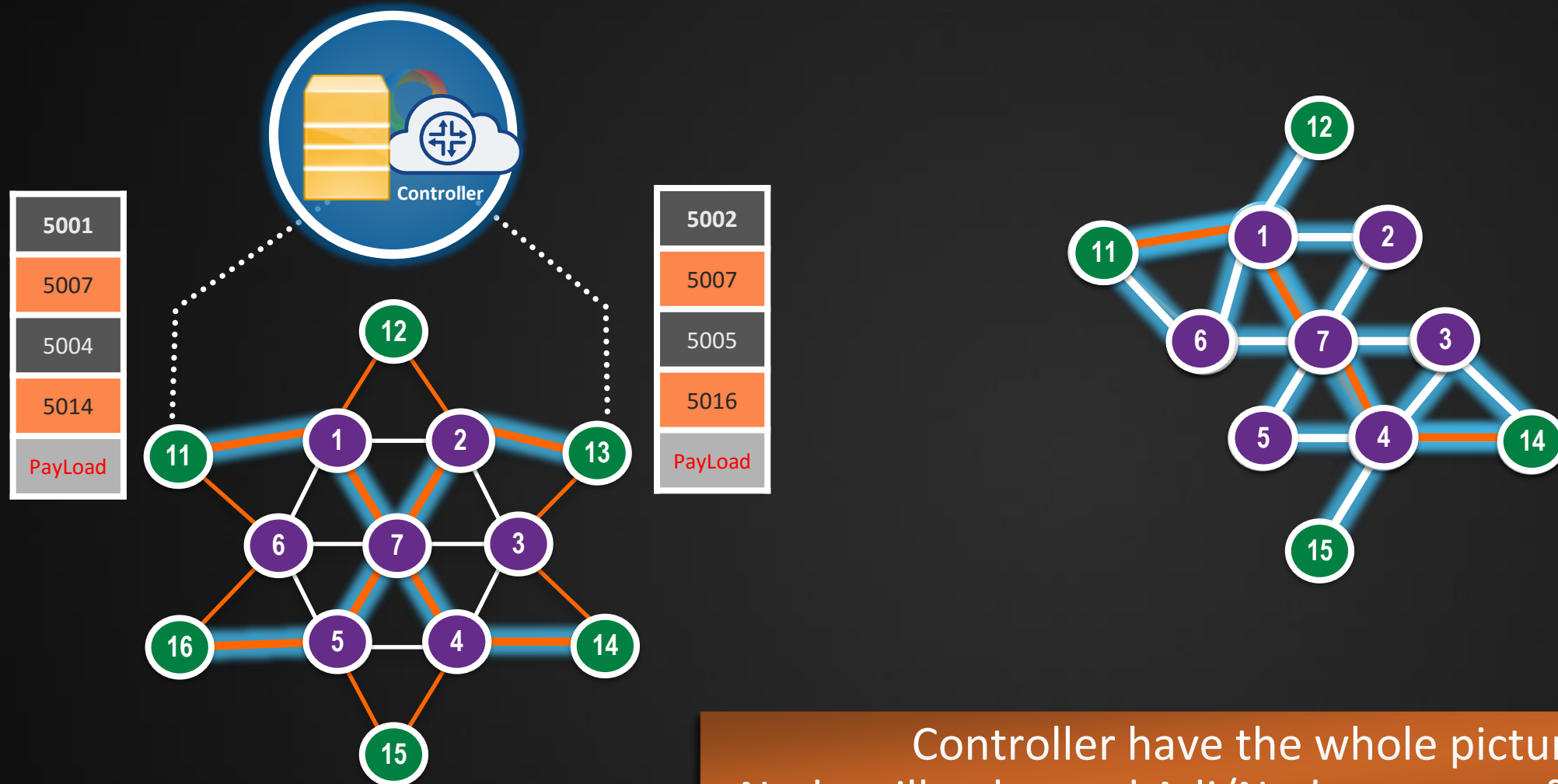
Segment Routing Architecture

Step1: Build SR Topology by IGP Ext Advertisement



Segment Routing Architecture

Step2: Controller calculate/program Label stacks from Edge

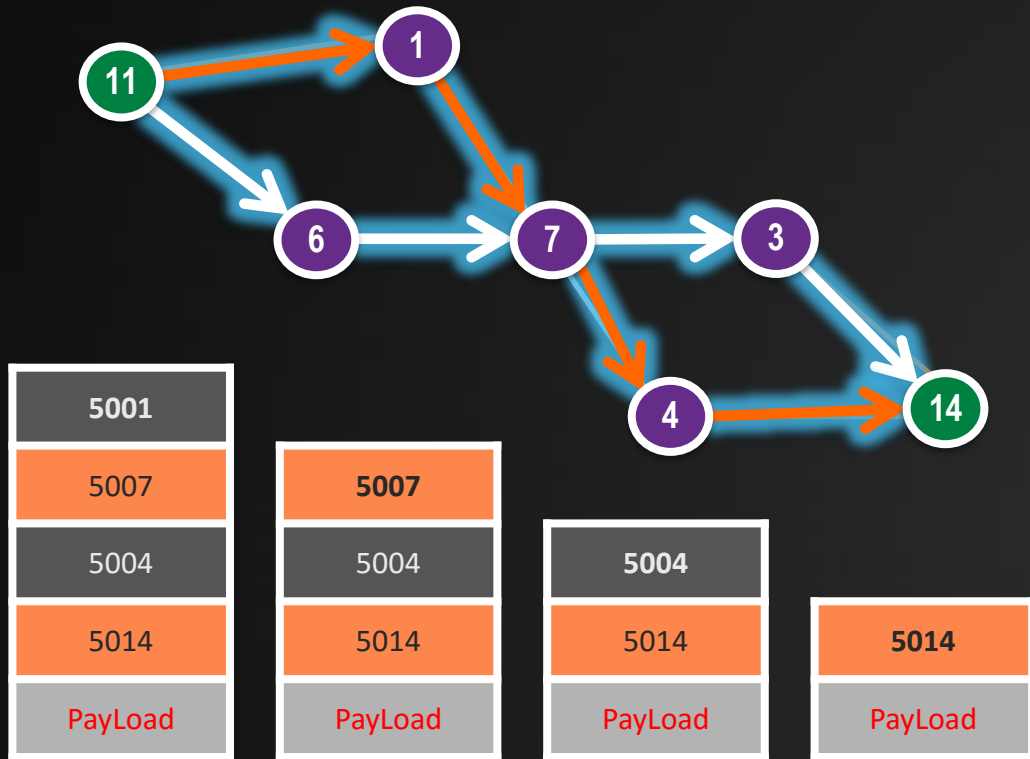


Controller have the whole picture
Node will only need Adj/Node segment forwarding

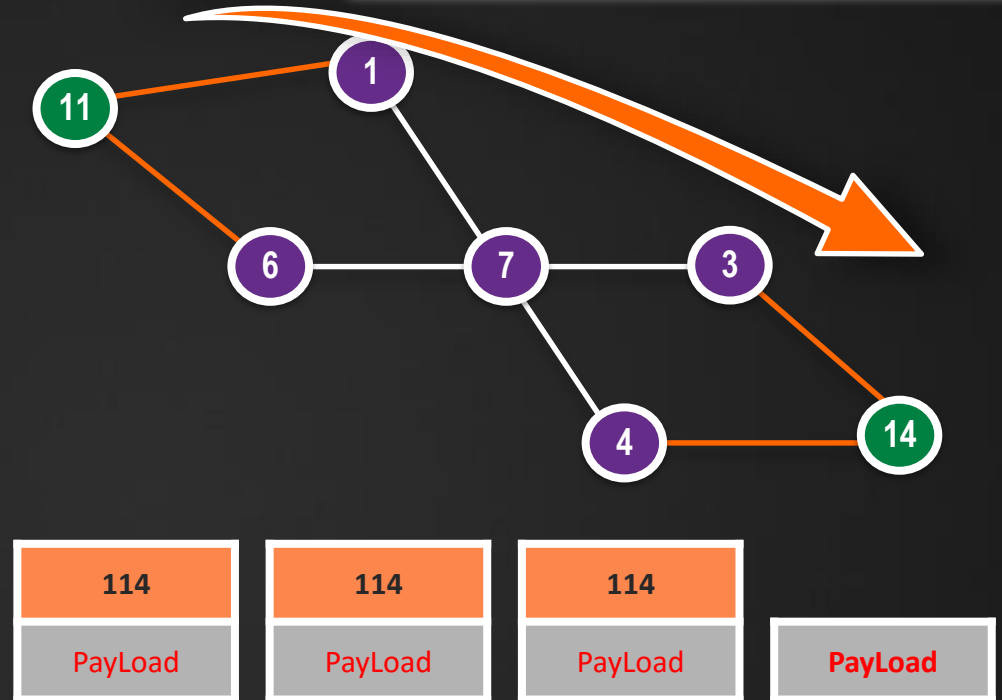
Adj/Nodal Segment forwarding

Nodal/Adj Label space is different, No Recursive look up.

Packet injected anywhere with label 114 will reach node 14



- Node Advertise Adj label, IGP extension
- Only install Adj label on router, not aware of rest network.
- Push multiple labels stack to reach remote router
- POP label only

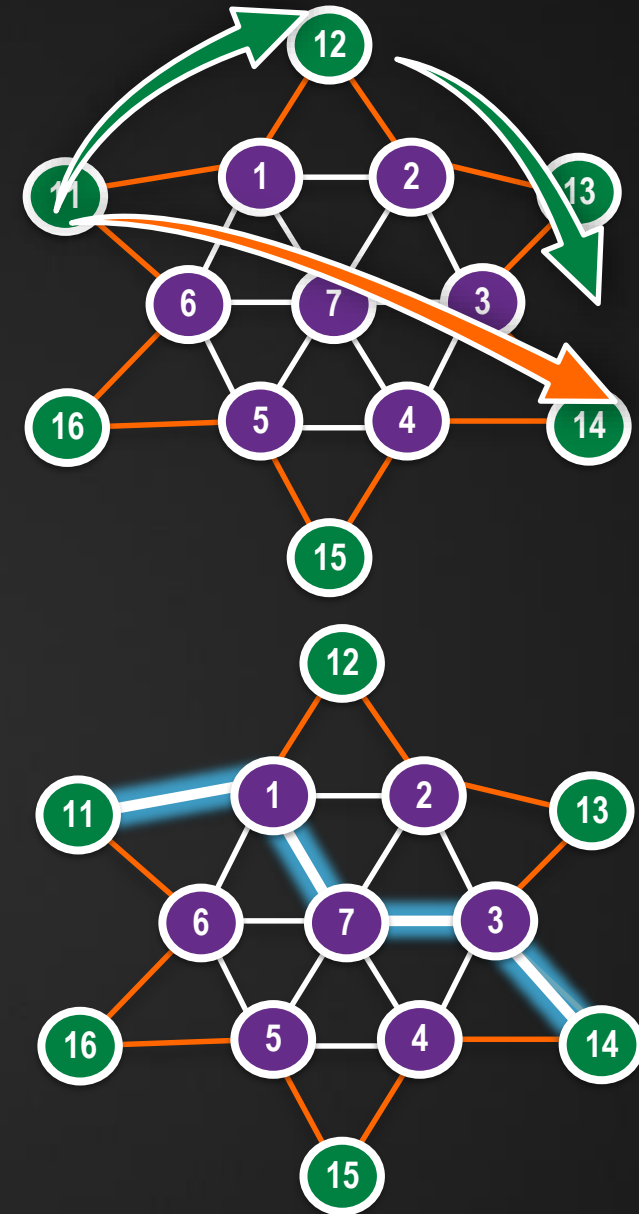


- Node advertise, unique {64-5000}
- IGP extension, normal SPF for all loopback
- Nodal label keep same in every nodes
- Swap Label Only

Path Creation

Source Based Routing

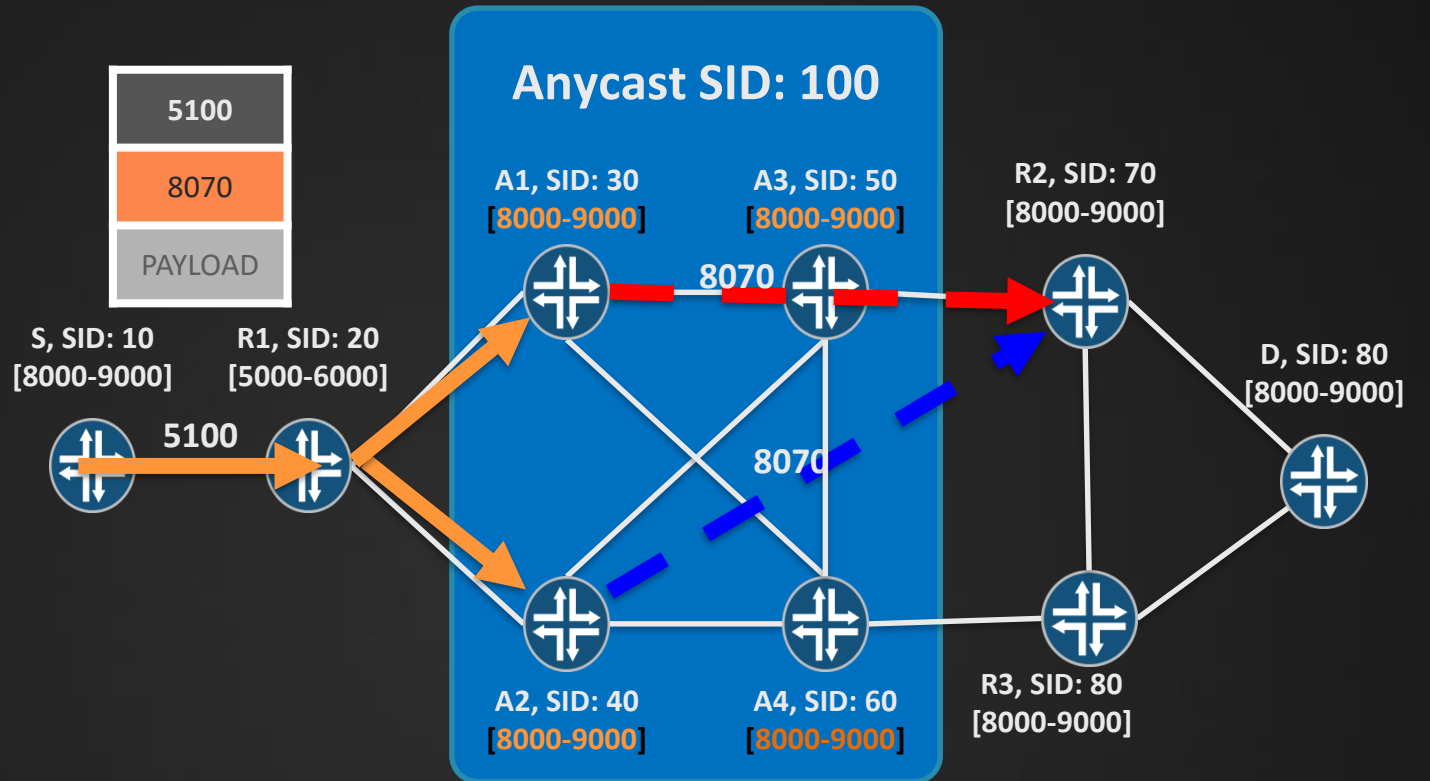
- A. Follow the IGP
 - one label pushed, the nodal segment(Node-SID),
 - SPF can leverage the ECMP path
 - Example, {114}
- B. Explicit Via nodal (like loose node in RSVP-TE)
 - Push list of via nodal...
 - Between nodal, SPF load balance.
 - Easy to expanded across Area/AS
 - Example, {112,114}
- C. Explicit via Adj, any path
 - Push of list of Via Adj
 - Example, {5001,5002,5003,5004,114}
- D. Mixed Path with Adj/Nodal



ANYCAST SEGMENT ID FOR NODE REDUNDANCY

draft-psarkar-spring-mpls-anycast-segments-01

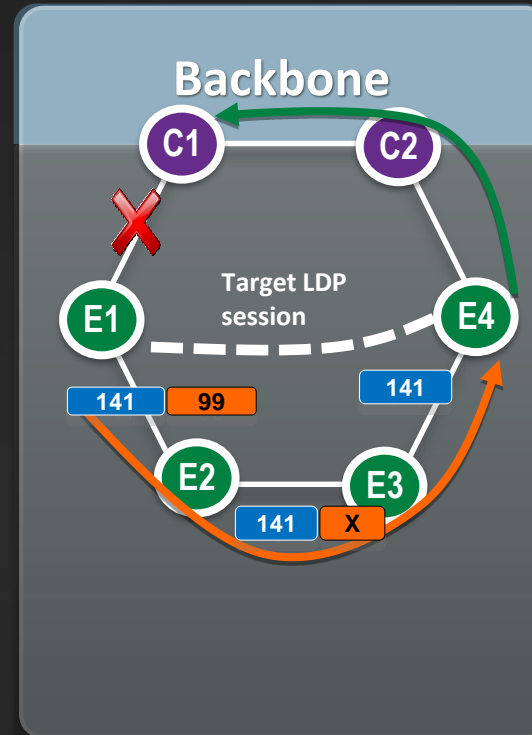
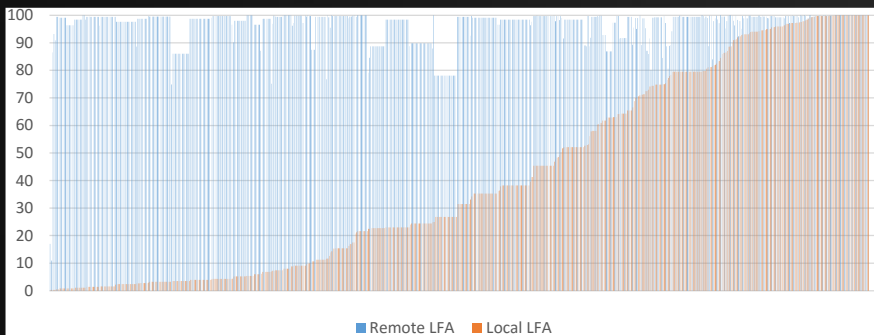
- Anycast SID
 - A group of Nodes share the same SID
 - Work as a “Single” router, single Label
- Any Topology
 - Hub/Spoke
 - Ring Topology
 - Anycast and other nodes follow IGP
- Application
 - ABR Protection
 - Seamless MPLS
 - ASBR inter-AS protection



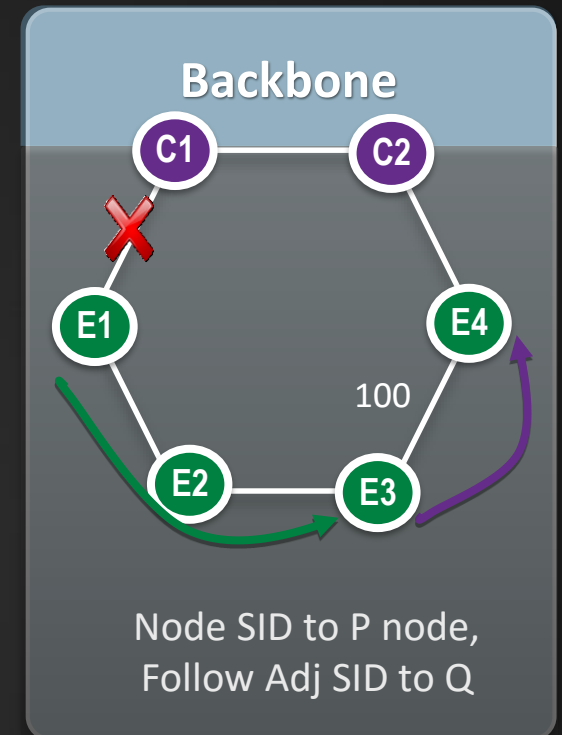
TI-FRR/TI-LFA

SEGMENT ROUTING CAN GUARANTEE 100%

- IP-based FRR not guaranteed in any topology
- Directed LFA (DLFA) is guaranteed when metrics only cover few cases, extra computation (RLFA) also 90%+ topology
- TI-FRR, Target LDP session with RSVP Tunnel
- TI-LFA Segment Routing, 2 actions
 - node segment to P node(From E1, can reach C1 without via failure link.
 - adjacency segment from P to Q Node(From Q node can reach C1 without via failure Link)
 - TI-LFA 100% Guarantee



IP FRR



Segment Routing FRR

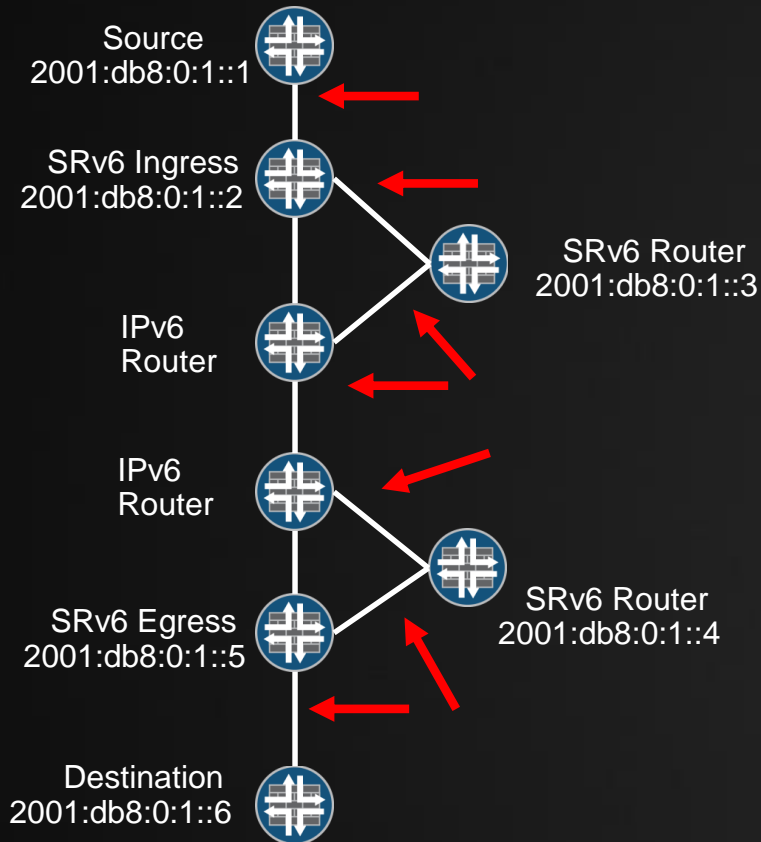


SRV6 STANDARDIZATION

- IETF is in the process of standardizing SRv6
 - Draft-ietf-6man-segment-routing-header-01
 - Work in Progress
- Two modes of operation
 - Insertion mode
 - SR ingress router inserts an SRH between IPv6 header and IPv6 payload
 - SR egress router optionally removes the SRH
 - Prepending mode
 - SR ingress router prepends a new IPv6 header and an SRH to the original IPv6 header
 - SR egress router always removes the new IPv6 header and the SRH, leaving only the original IPv6 header

Segment Routing IPv6(Animated)

include a SRH, Insertion mode and Prepending mode



▪Draft-ietf-6man-segment-routing-header-01

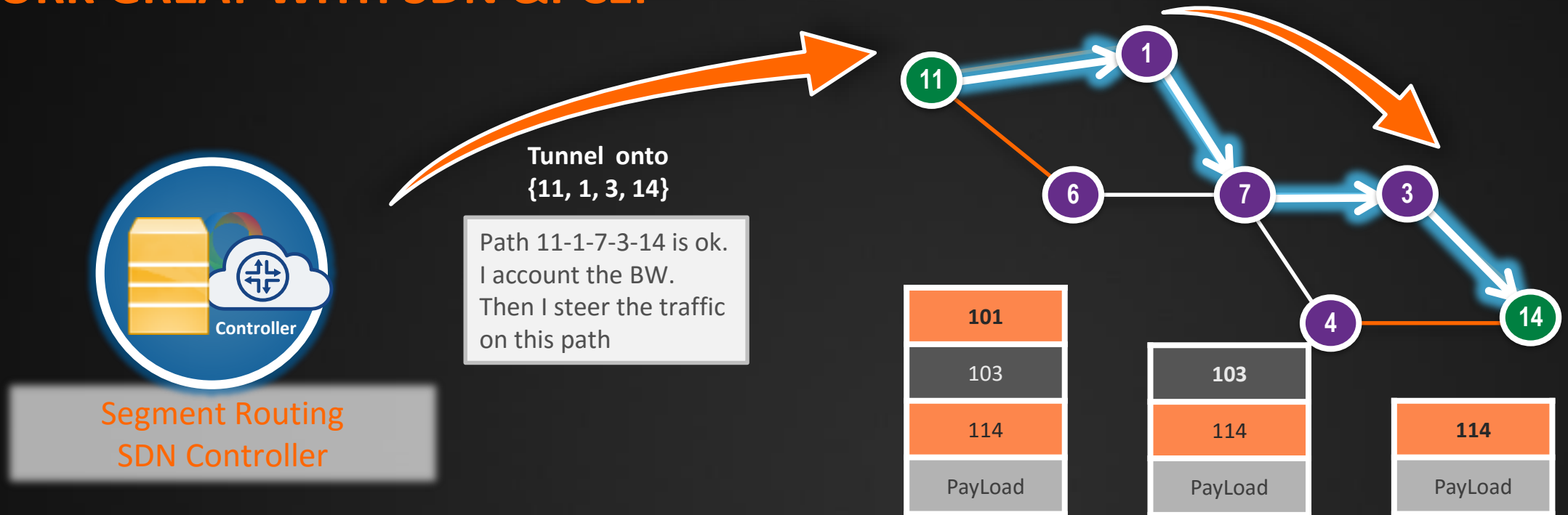
Ver	DSCP	Flow Label	
Length 136		Next HDR SRH	Hop Limit 255
Source Address 2001:db8:0:1::1			
Destination Address 2001:db8:0:1::6			
Next HDR TCP	Length 56	HDR Type 4	Seg Left 2
First Seg 2	Flags C = 1		Reserved
Segment 0 2001:db8:0:1::6			
Segment 1 2001:db8:0:1::5			
Segment 2 2001:db8:0:1::4			

IPv6
HEADER

~~Segment~~
Routing Header

TCP Header

SEGMENT ROUTING SDN WORK GREAT WITH SDN & PCEP



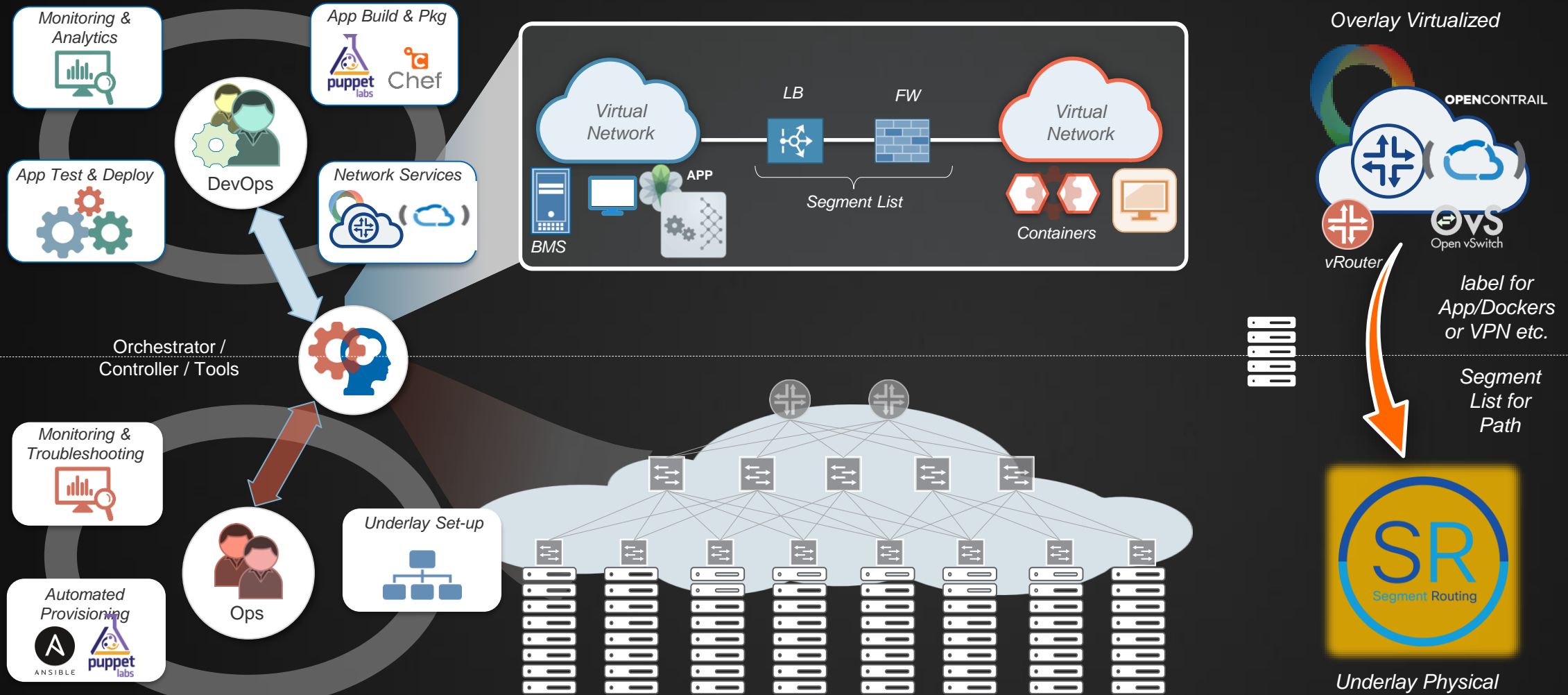
- The network is simple, highly programmable and responsive to rapid changes
- Source Based routing, label pushed in the source will decide the path.
- On router, PCE Client **no need signaling protocol to create path**, Just Segment Routing.
- Better than PCE+RSVP-TE, No on-demand signaling the path.
- Better than Static MPLS label push from SDN, SR still have ECMP, Resilience, FRR.

Segment Routing vs MPLS

Features	MPLS	Segment Routing
Control Protocol	LDP/RSVP/BGP(any of label allocation) OSPF/ISIS, BGP (any of topology), SDN	OSPF or ISIS or BGP, or SDN Controller
Traffic Engineer	RSVP, PCE Client, SDN	OSPF/ISIS(option) SDN (option)
Fast Reroute	LDP FRR, or RSVP-TE FRR	Build in FRR, cover for all scenario
Inter-Area/Inter-AS	With help of BGP label, or RSVP-TE inter Area hard to protect	Loose Node ID extension
Source Path Routing	No, IGP only	Yes, explicit indicate ingress
Scalabilities	LDP same as IGP....RSVP limited.	Node + ADJ Segment(less entry) Best Scale
Performance Measurement	NO	Build in with RFC 6374
SDN integration	PCE, RSVP-TE	PCE, BGP-LU, SR

SEGMENT ROUTING FOR CLOUD DEPLOYMENT

UNDERLAY PATH BY SR PROTOCOL, OVERLAY SDN CONTROLLER WITH LABEL APP



AGENDA





Introduction

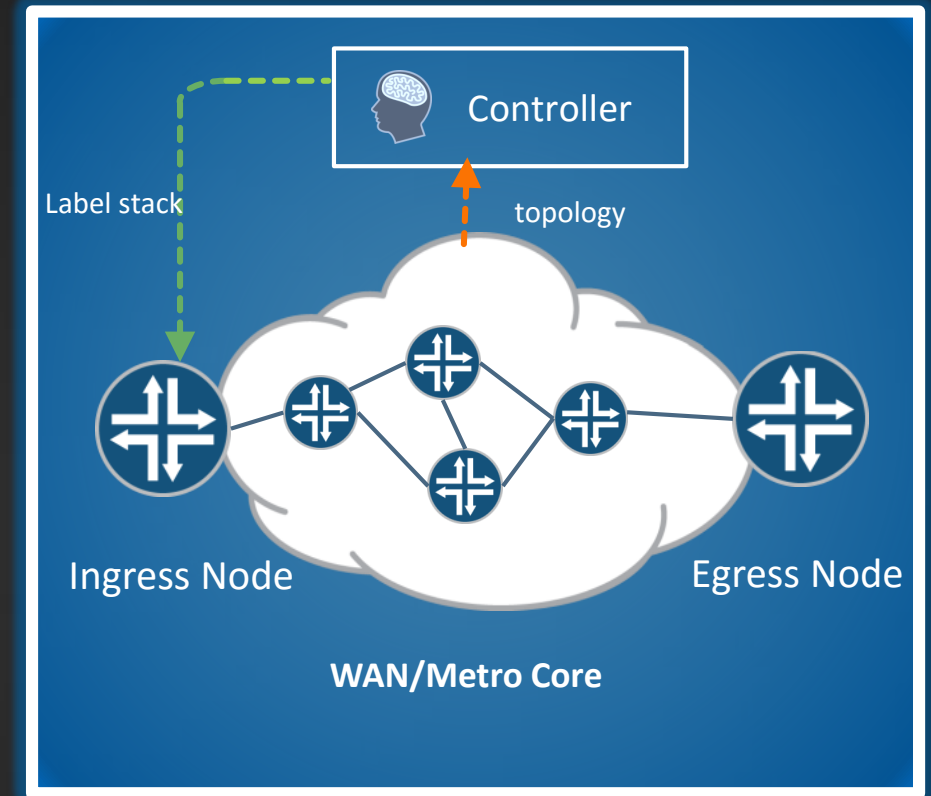
Segment Routing Deep Dive

Segment Routing SDN and Use Case

Summary

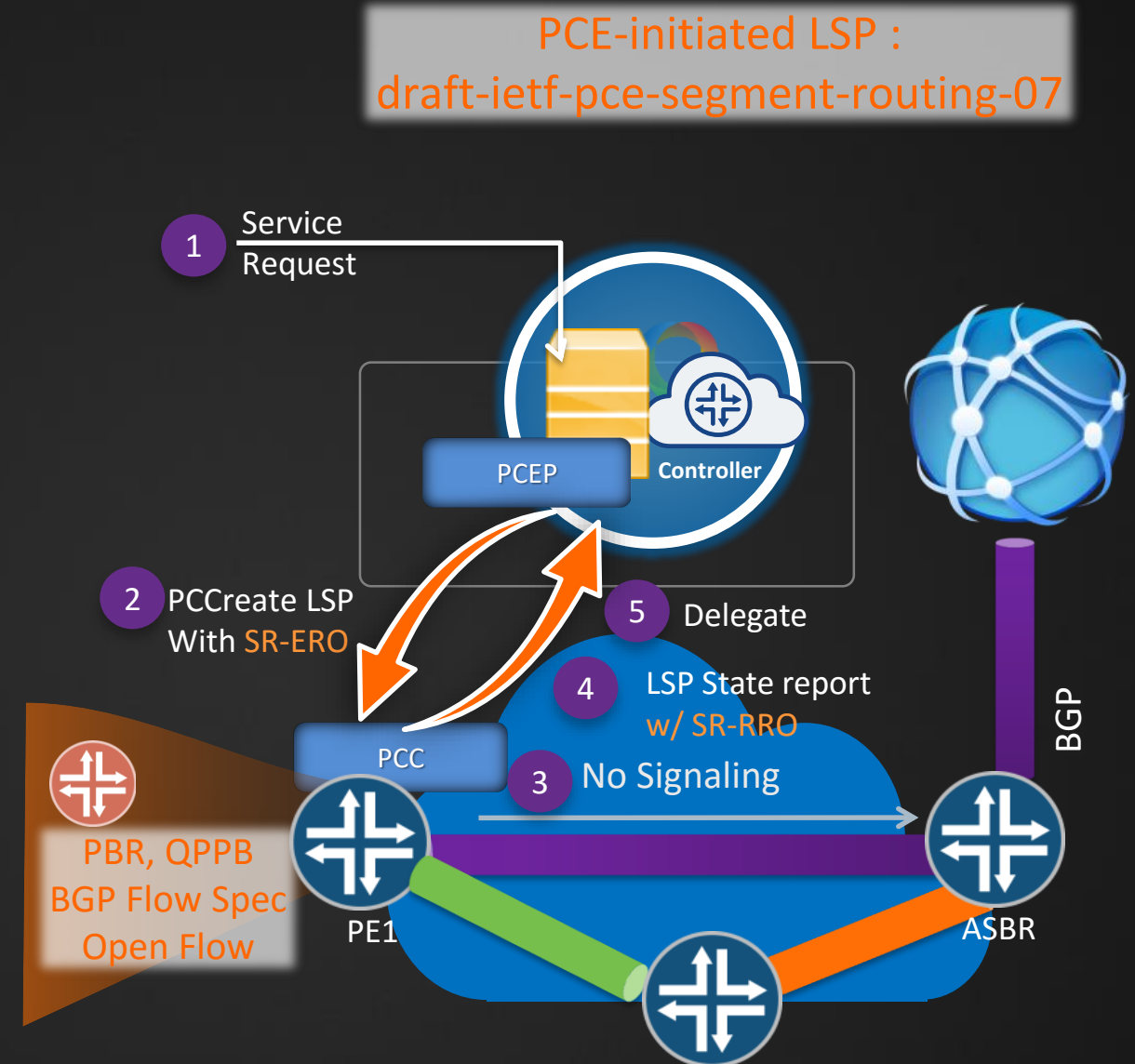
SPRING : DOMAIN APPLICABILITY

	Data Center	Fixed design, EBGW as IGP, Simpler mgmt. with common SRGB
	WAN	Alternate way of doing FRR, No core state, BGP-LS to export topology to controller
	Metro	FRR in Metro rings, PW transport
	Edge	Traffic engineering, Northbound interface: PCEP, BGP-LU, Flow-spec



PCE WITH SEGMENT ROUTING

- PCEP SR similar with RSVP-TE PCEP
 - Open message negotiate SR-PCE-CAPABILITY TLV
 - PCCreate LSP with SR-ERO for Label stack
 - No Need Signaling on PE-P-PE
 - LSP State report with SR-RRO
- BGP-LS get the network information
 - TEDB information with label send back to Controller
 - draft-gredler-idr-bgp-ls-segment-routing-ext-xx.txt
- Service mapping by
 - Openflow/PBR/QPPB/BGP FlowSpec



BGP FlowSpec redirect to SR LSP Tunnel

Type	Matching	Type	Matching
Type 1	Destination prefix	Type 7	ICMP type
Type 2	Source prefix	Type 8	ICMP code
Type 3	IP protocol	Type 9	TCP flag
Type 4	Port (Defines a list of pairs that matches source or destination UDP/TCP ports)	Type 10	Packet length
Type 5	Destination port	Type 11	DSCP
Type 6	Source port	Type 12	Fragment

Type	Extended Community	Encoding
0x8006	Traffic-rate	2 byte/4 byte float
0x8007	Traffic-Action	bitmask
0x8008	Redirection	6-byte route-target
0x8009	Traffic-marking	DSCP Value

NOTE: Detailed information about each type and field can be found in RFC 5575 section#4 “Dissemination of Information”.

Segment Routing with PCEP and BGP-LS

- Prefix & node SID learning via ISIS &/or BGP-LS
- New PCEP capability, ERO subobject and TLVs
 - ✓ draft-ietf-pce-segment-routing-06
- SPRING-TE LSP creation, visualization & optimization

The screenshot shows the Juniper NorthStar Controller interface. On the left, a timeline displays recent events related to LSP provisioning. The main area features a network topology diagram with nodes and links. Below the diagram is a table listing the configured LSPs.

Name	Node A	Node Z	Planned Bandwidth	Control Type	Op Status	Explicit Route	Label Stack	IP Z
SPRING-LSP1	vmx101	vmx104-11	500M	PCEInitiated	Up	11.0.0.101 - 11.101.105.2 - 11.105.107.2 - 11.114.117.1	11.0.0.101 - 299776 - 300064 - 299872	11.0.0.1...
SPRING_DIVA	vmx101	vmx103-11	100M	PCEInitiated	Up	11.0.0.101 - 11.101.105.2 - 11.105.107.2 - 11.103.107.1	11.0.0.101 - 299776 - 300064 - 299840	11.0.0.1...
SPRING_DIVB	vmx10...	vmx104-11	100M	PCEInitiated	Up	11.0.0.102 - 11.102.106.2 - 11.104.106.1	11.0.0.102 - 299840 - 299840	11.0.0.1...

The screenshot shows the Juniper NorthStar Controller interface with a 'Provision Diverse LSP' dialog box open. The dialog allows configuring two tunnels for a diverse LSP. The configuration includes names, nodes, bandwidth, and diversity options.

Provision Diverse LSP

Properties | **Scheduling**

Tunnel 1

- Name: SPRING_DIV_A
- Node A: vmx101
- Node Z: vmx103-11
- Bandwidth: 100M
- Coloring: []
- Setup: 7
- Hold: 7
- Comment: []

Tunnel 2

- Name: SPRING_DIV_A
- Node A: vmx102-11
- Node Z: vmx104-11
- Bandwidth: 100M
- Coloring: []
- Setup: 7
- Hold: 7
- Comment: []

Diversity

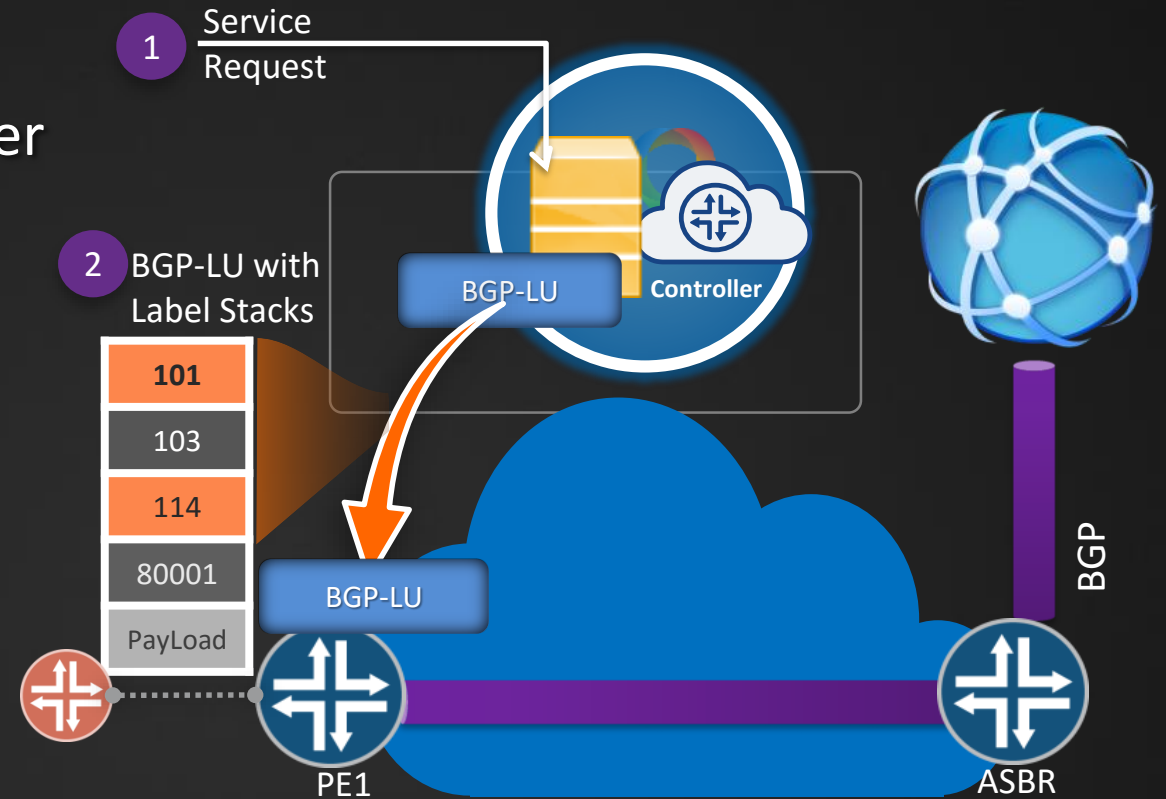
Link Site SRLG

Preview Paths | **Cancel** | **Submit**

BGP-LU WITH SEGMENT ROUTING

draft-rosen-idr-rfc3107bis-00.txt
NOT draft-ietf-idr-bgp-prefix-sid-03

- BGP-LU Session between Controller/Router
 - BGP LU carrier the label stack for SR/LSP
 - BGP-LU carrier the Label stack for LSP + VPN Service
- BGP-LS get the network information
 - TEDB information with label send back to Controller
 - draft-gredler-idr-bgp-ls-segment-routing-ext-xx.txt
- BGP is the only protocol for Service and Tunnel
 - QPPB/BGP FlowSpec
 - With additional Openflow/PBR



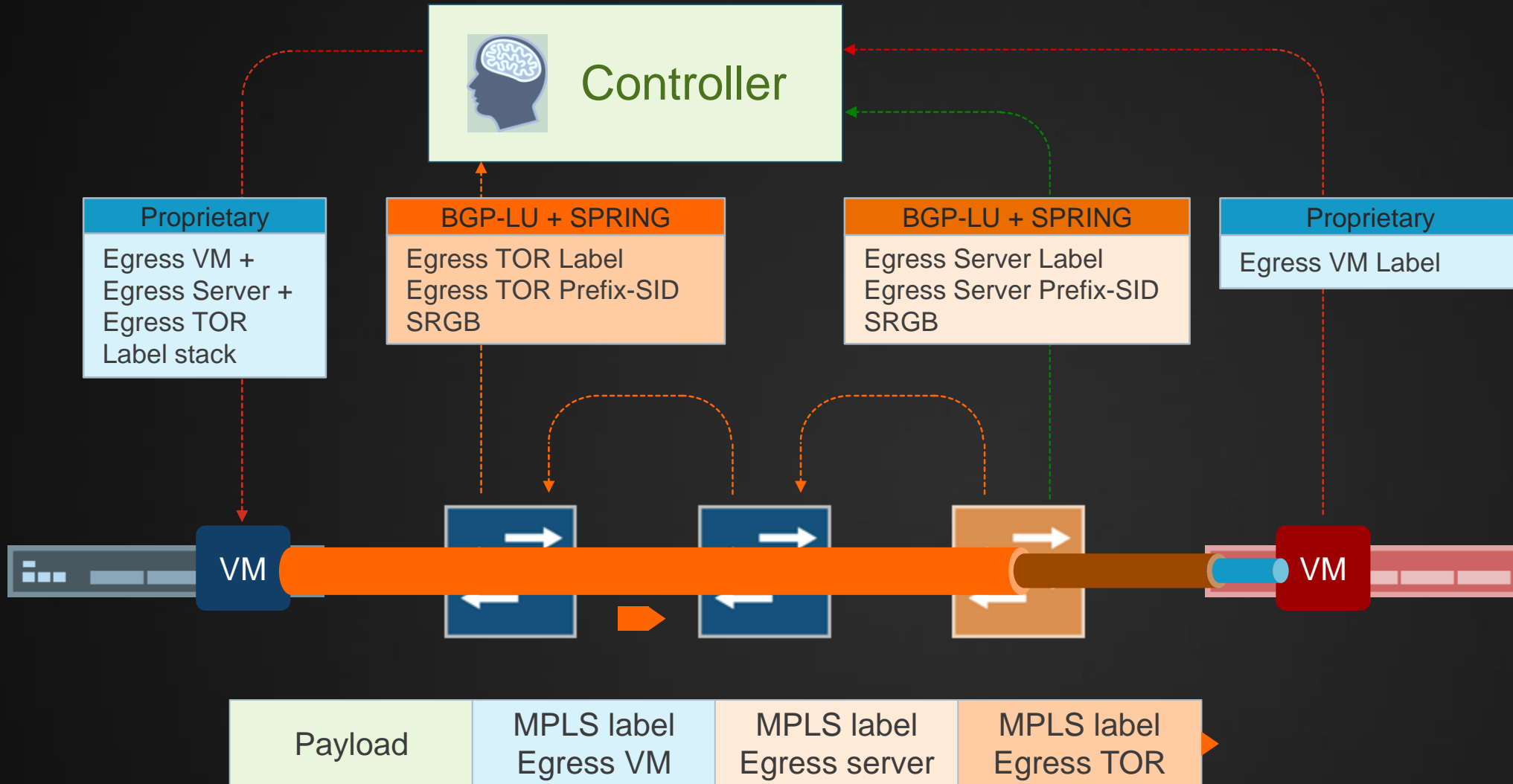
Example from ExaBGP

```
bespalov@CentOS-1 ~/exabgp-3.4.16/sbin>cat ~/bespalov/config/exabgp neighbor 192.168.255.12 {
local-address 192.168.255.2;
peer-as 65000; local-as 65000;
family { ipv4 nlri-mps; }
static {
    route 10.255.255.8/32 {
        next-hop 10.0.0.2;
        label [ 800005 800007 800006 800008 ]; }}
}
```

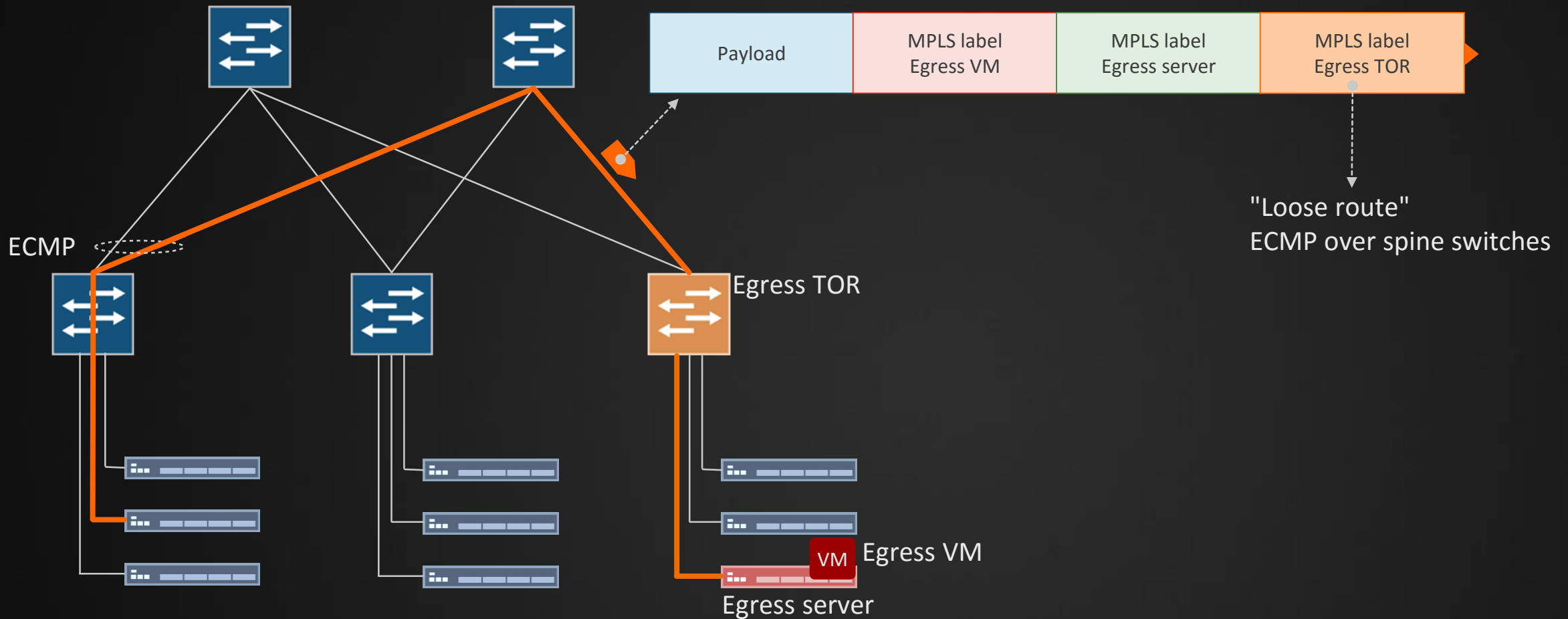
MPLS IN DATA CENTERS

- Overlays are widely used today
 - South → North: Egress Peer Engineering (EPE)
 - North → South: Load balancing, Floating IPs, ...
 - East ↔ West: Multi Tenancy
- Currently overlays are IP-based, moving to MPLS
 - Consistent end-to-end protocol; avoid 'impedance-mismatch' at boundaries
 - Hierarchical Forwarding [MPLS Label Stack]; reduces FIB state
- Use SPRING-like approach
 - Label stacking (hierarchy) to reduce FIB size on switches with merchant silicon
 - Label stacking for 'source-routing' across WAN
 - Different control plane inside data-center: BGP instead of IGP

SPRING INTRA DATA CENTER ROUTING

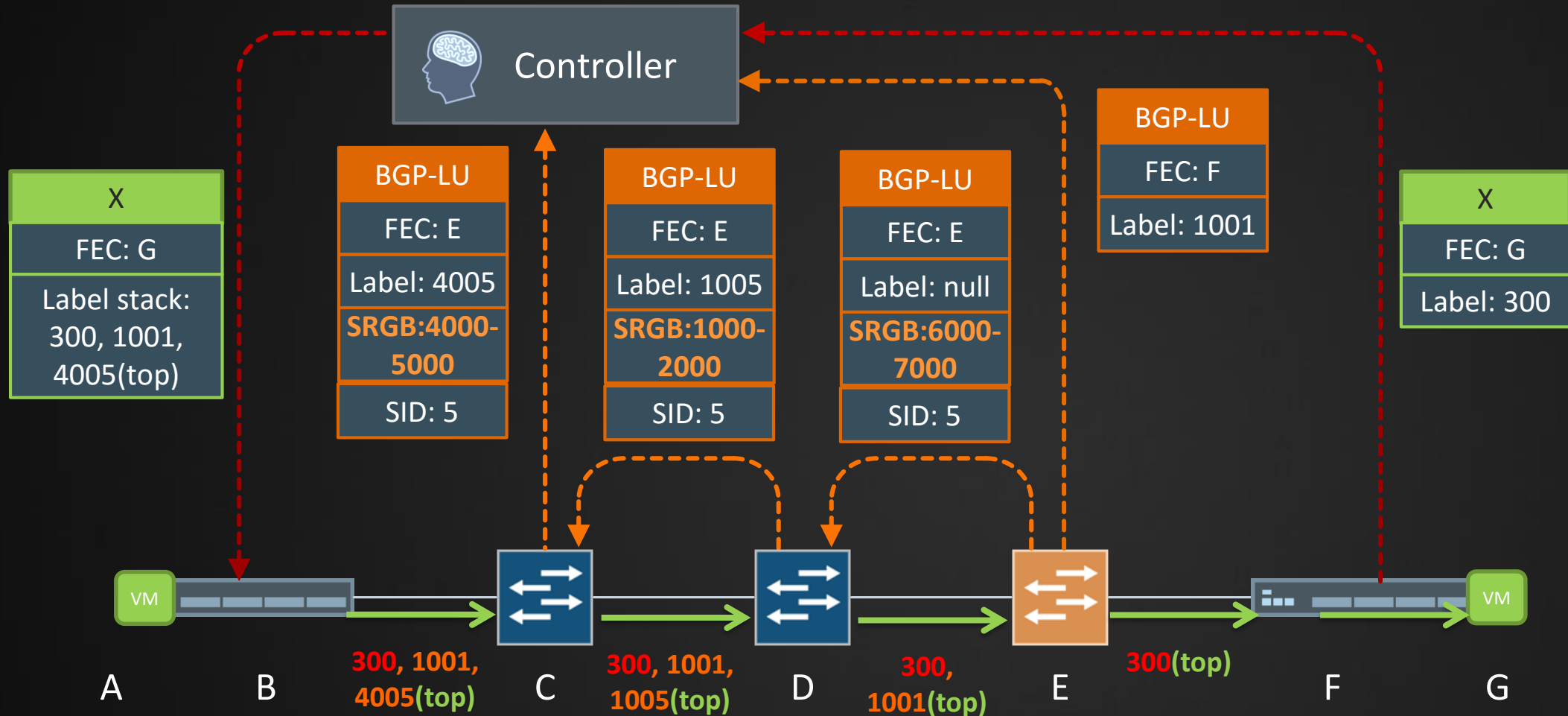


SPRING INTRA DATA CENTER ROUTING

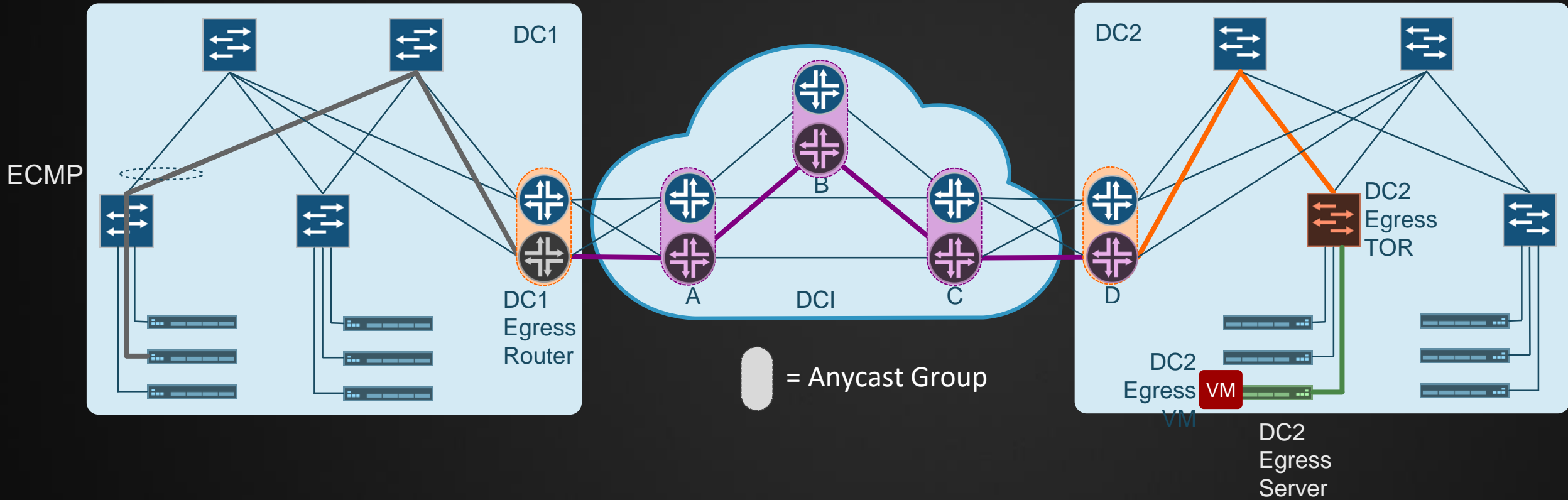
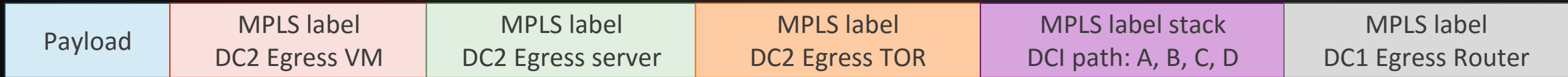


BGP-LU PREFIX SEGMENT PROPOSAL

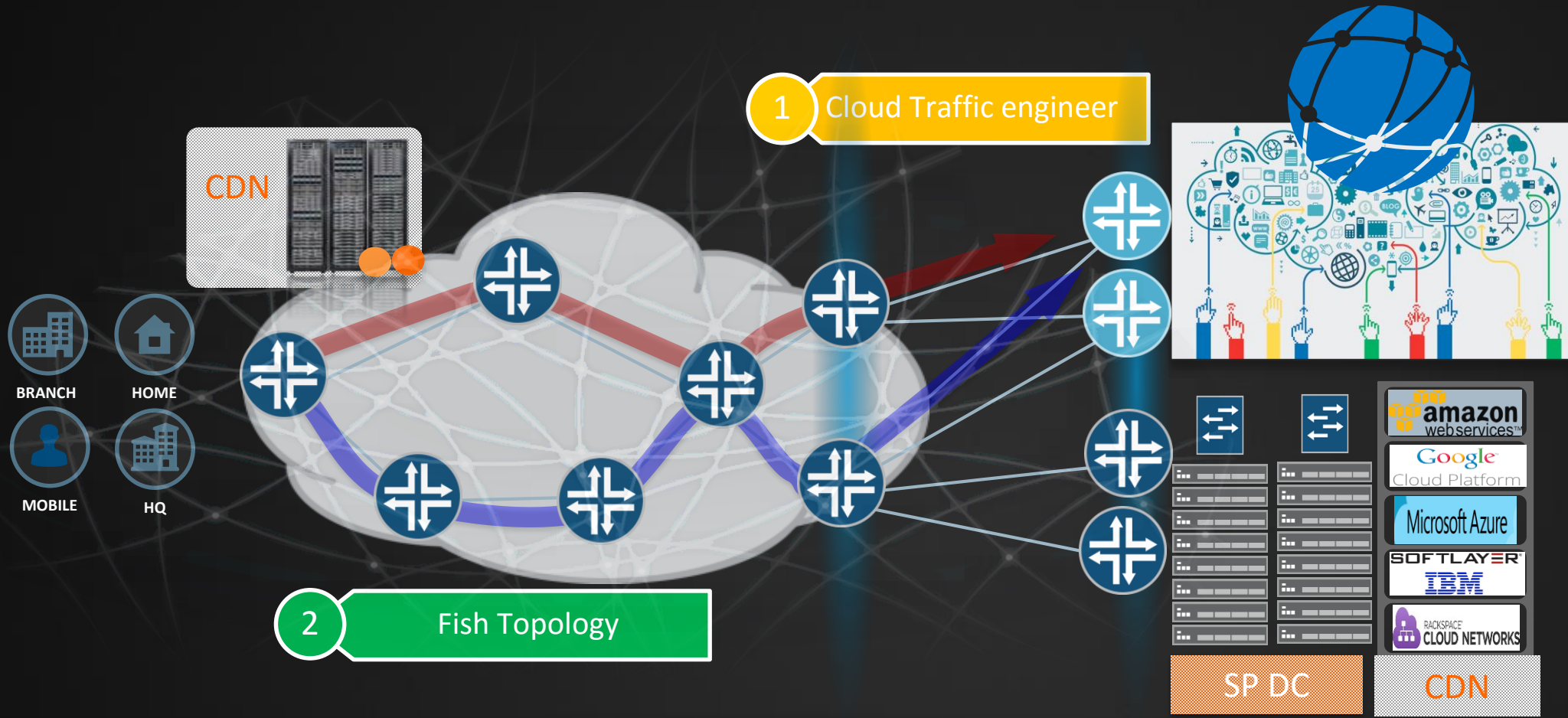
Juniper Proposal [draft-gredler-idr-bgplu-prefix-sid-00]



SPRING INTER DATA CENTER ROUTING



SPRING INTER-DOMAIN CLOUD TRAFFIC ENGINEER



Easy to optimize End-To-End Traffic for SP Owned Network.
How to optimize VIP Customer for Internet/Cloud connection?

BGP EPE DESIGN PHILOSOPHY

How to Select Which Peer to send

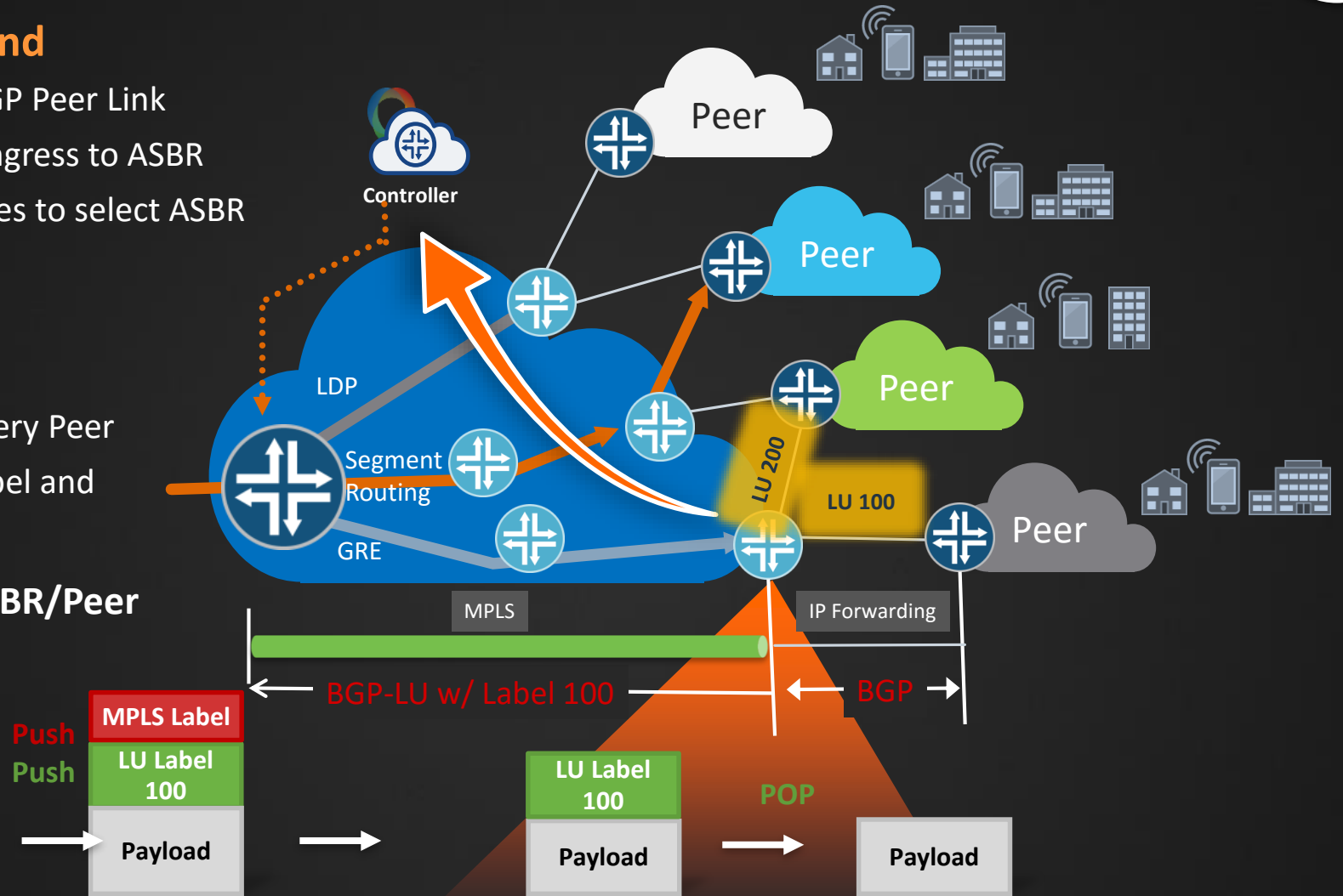
- Controller/RR may monitoring the BGP Peer Link
- Controller/RR find a tunnel from Ingress to ASBR
- Controller/RR based on certain rules to select ASBR

How ASBR identify a Peer

- Per Peer /32 address per label
- Install the MPLS Label POP for every Peer
- When ASBR received different label and send traffic to specific Peer

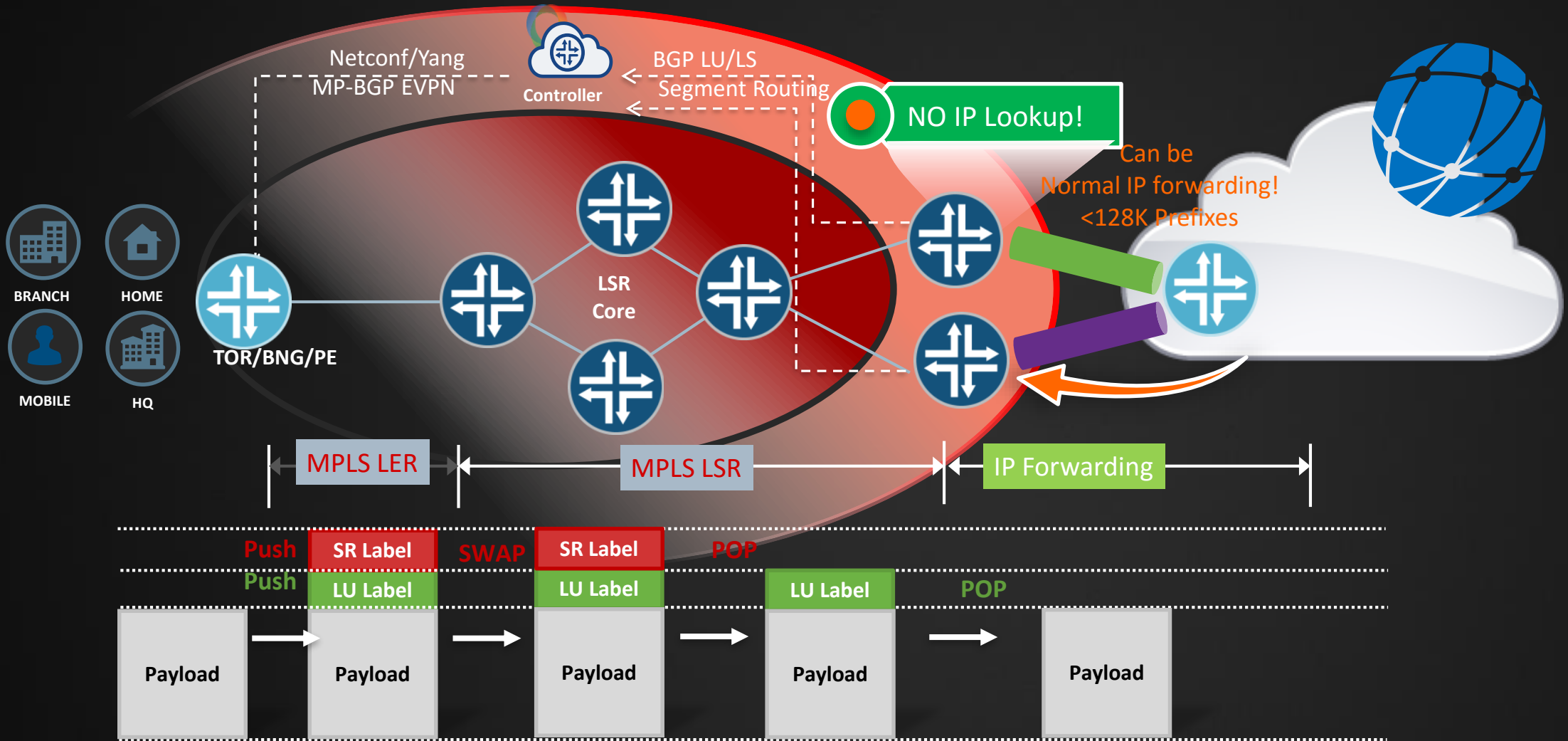
How Ingress mapping traffic to ASBR/Peer

- Ingress push tunnel label to ASBR
- Ingress push BGP-LU label

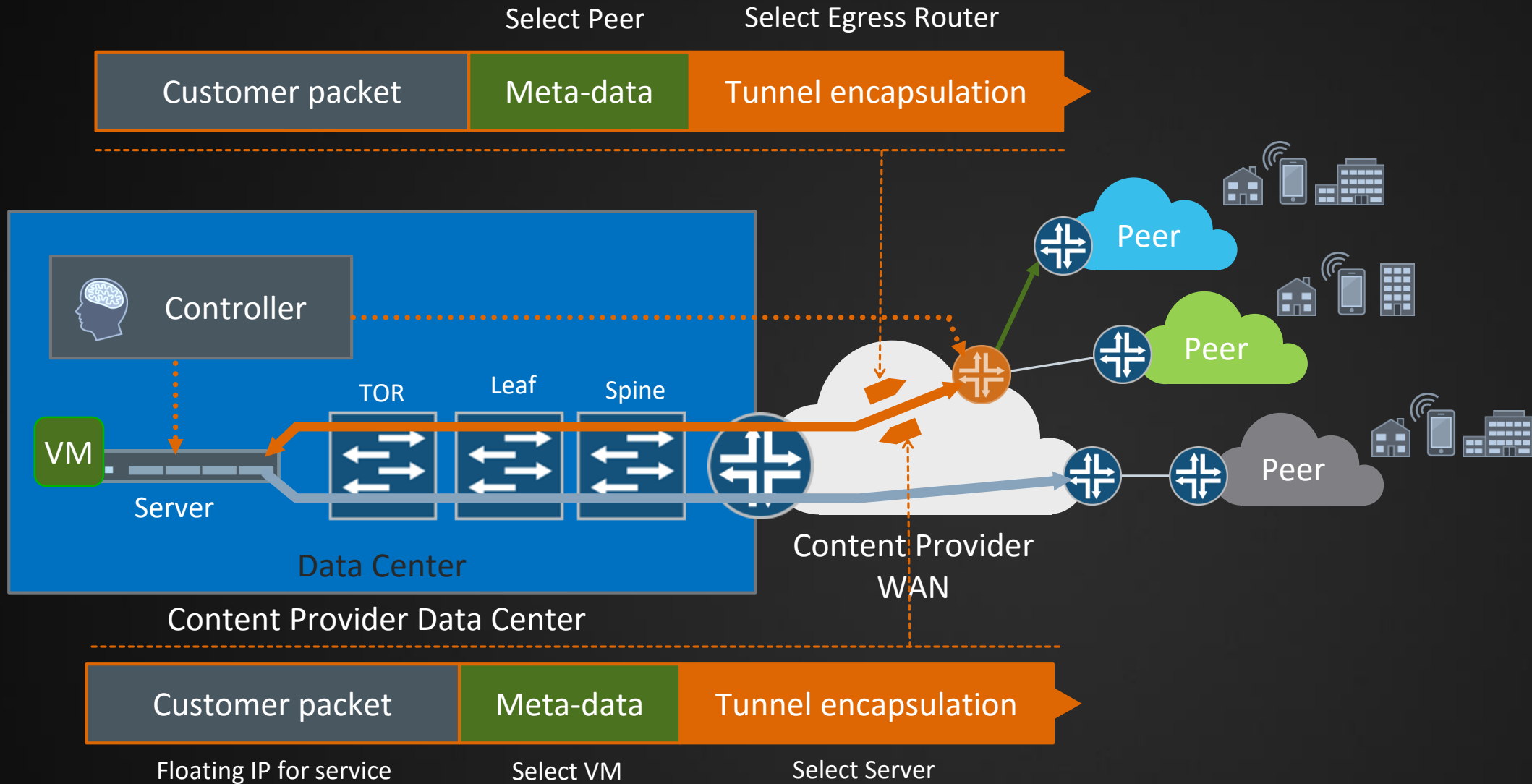


BGP-LU EPE & MPLS KEY BENEFITS

EXTEND HOLLOW CORE/LSR TO PEERING, CHEAPER PEERING SOLUTION

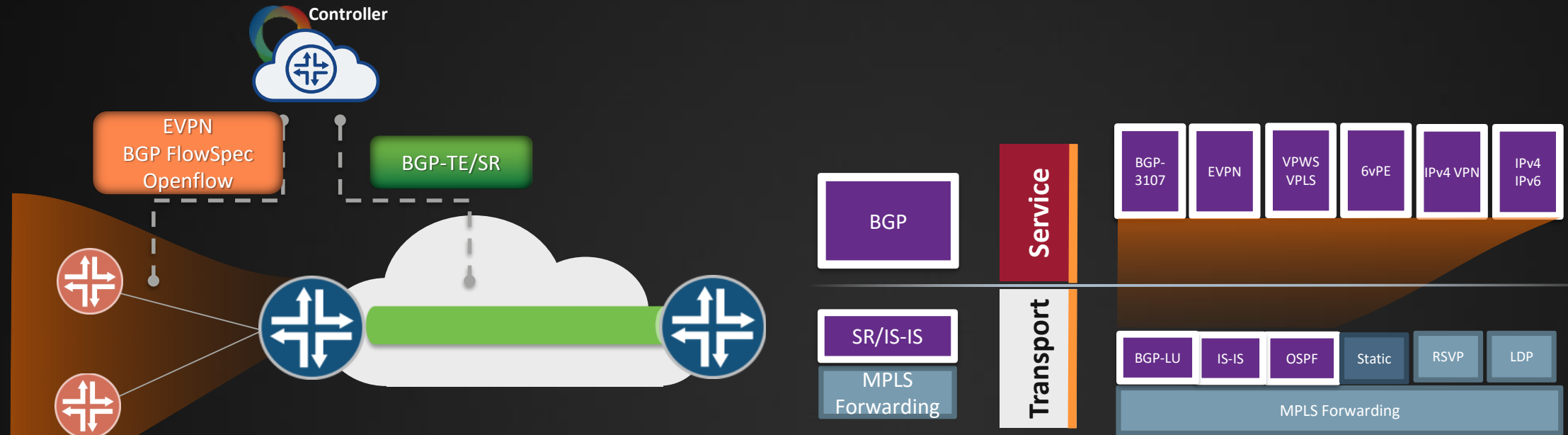


SEGMENT ROUTING AND EPE USE CASE



SEGMENT ROUTING IN ACCESS/AGGREGATION

SIMPLIFIED BOX FUNCTION, MOVE INTELLIGENCE TO CONTROLLER

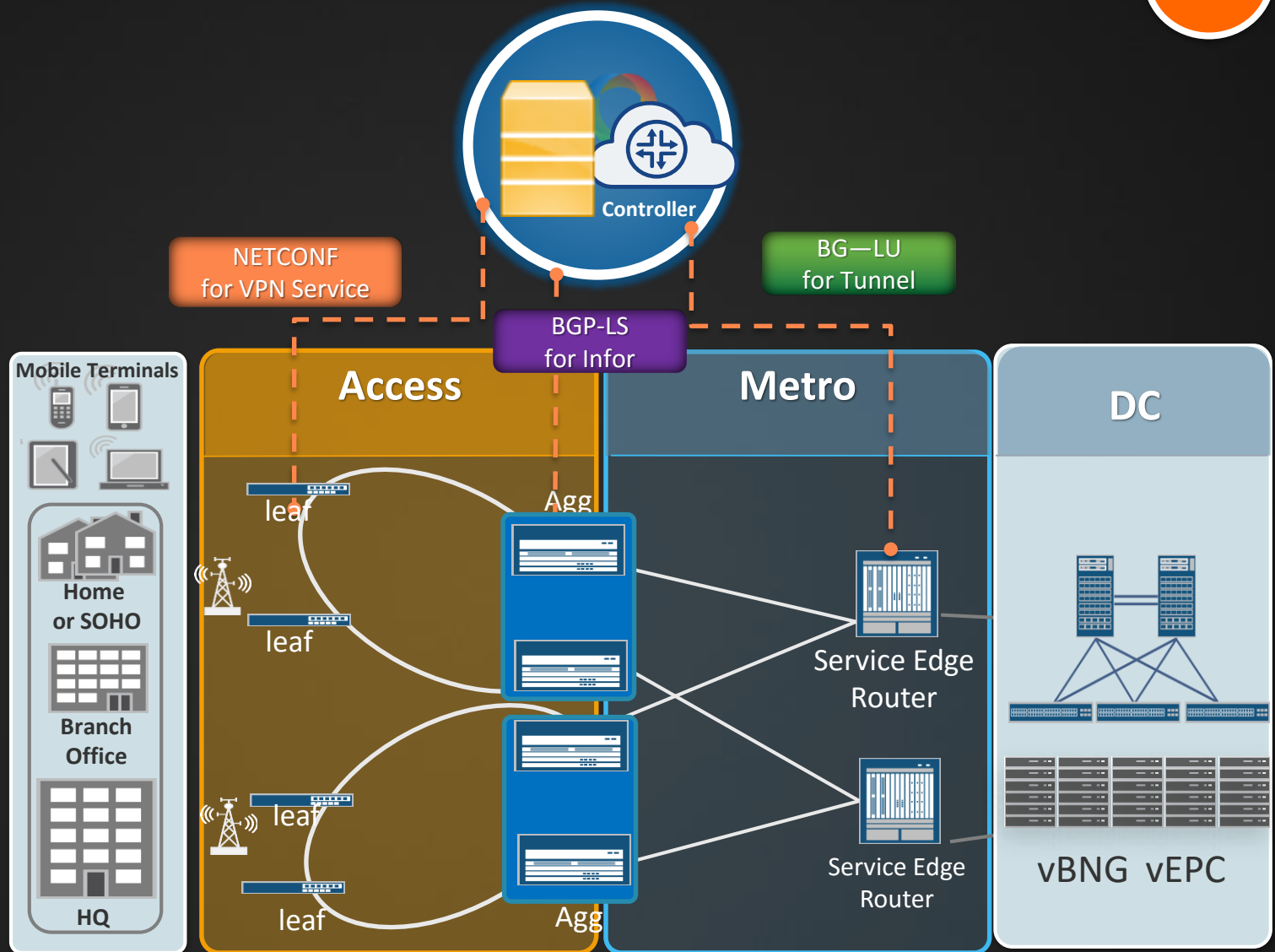


- Keep OAM/Clocking
- No need Peer with others, only Controller
- No Need Compute, Controller got full network view.

Minimal Protocols, Dumb Box in Access

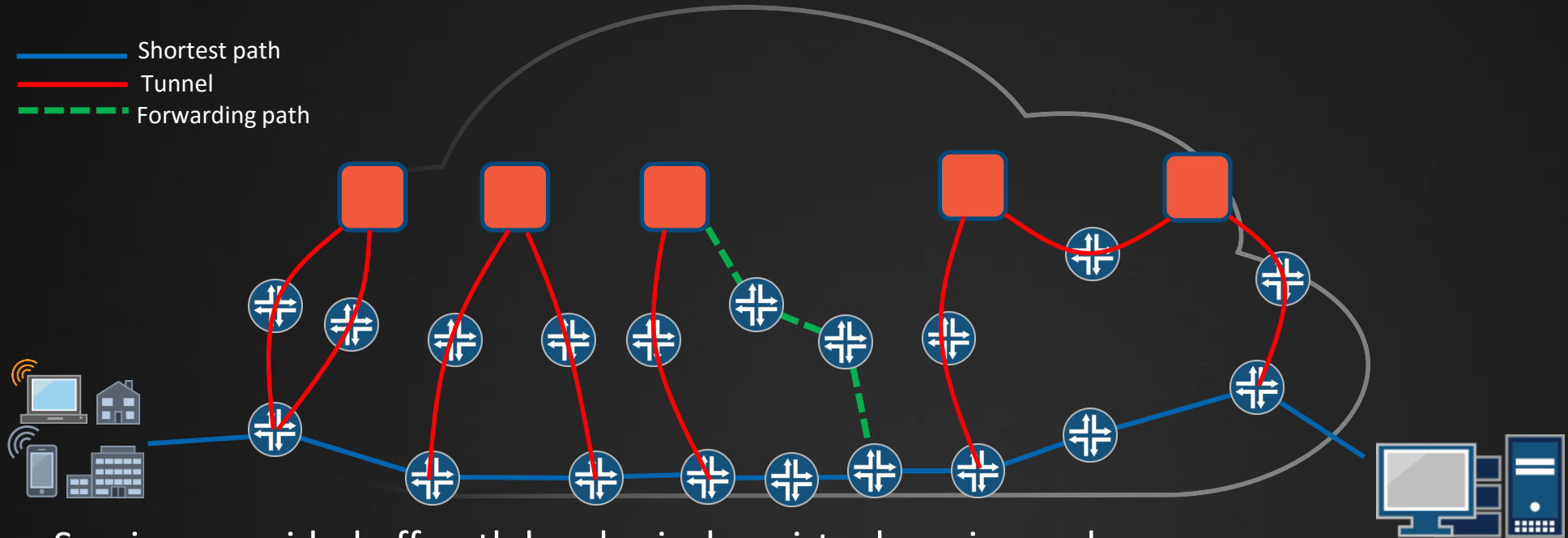
SEAMLESS MPLS EVOLUTION – SEGMENT ROUTING

- Architect Change
 - To manage 1,000+ boxes Add **SDN Controller**
 - RSVP-TE w/ RFC3107 to **Segment Routing**
- Technical Benefits
 - SP Fabric management with ZTP
 - Better **FRR** with LFA/RLFA/TI-LFA
 - Better ABR **Node protection** with Segment Routing Anycast SID
 - Better **tunnel provision** by BGP-LU or Controller
 - Better **Tunnel Stitching** by SR, no need RFC3107, save one label
 - Service Provision by NETCONF
 - Network information collect by BGP-LS



SEGMENT ROUTING FOR NFV SERVICE CHAINING

NO NEED NETWORK SERVICE HEADER(NSH), VNF SUPPORT MPLS



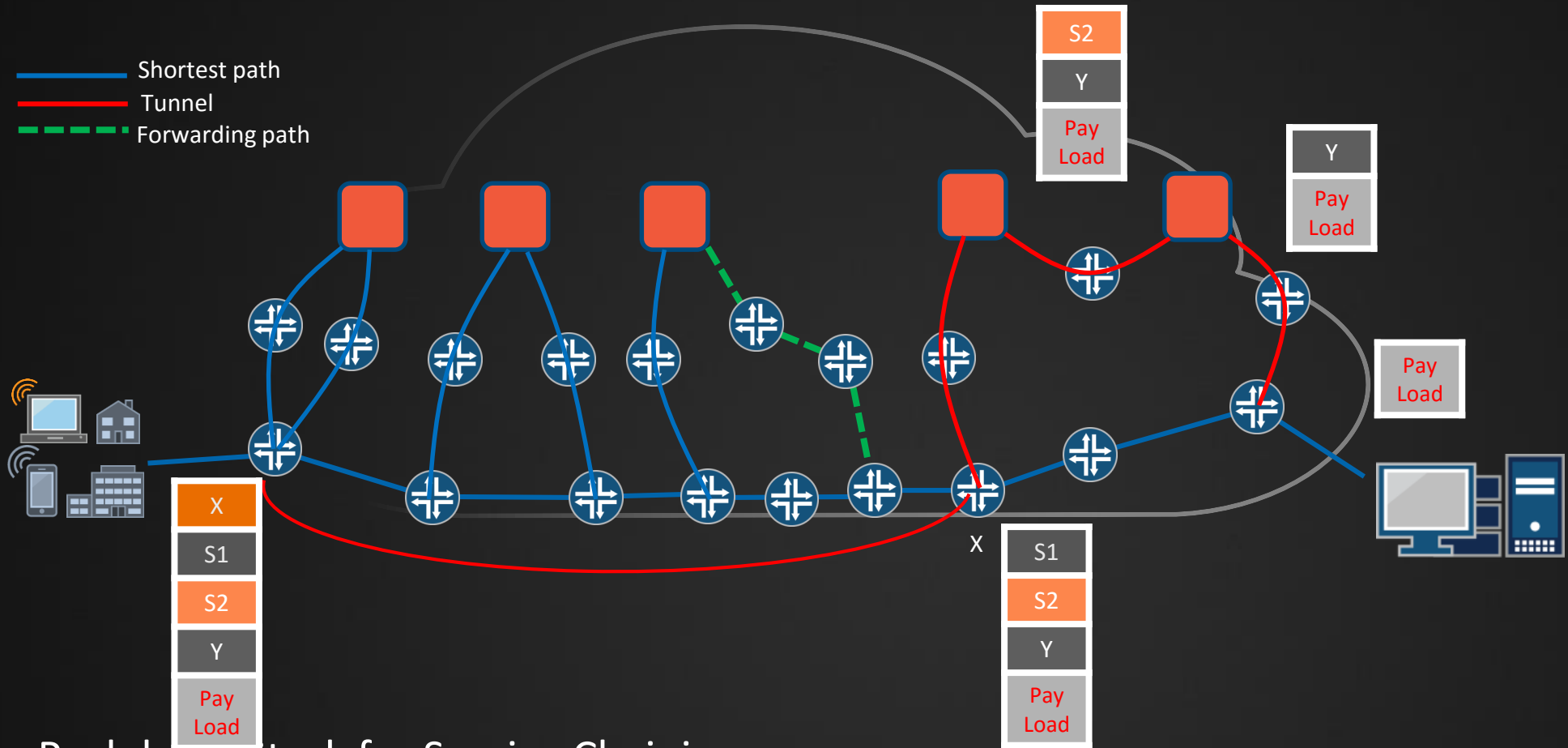
Services provided off-path by physical or virtual service nodes

Packets diverted through tunnels

- Return to forwarding path
 - By tunnel
 - Via forwarding
 - After attention by other service nodes

SEGMENT ROUTING FOR NFV SERVICE CHAINING

NO NEED NETWORK SERVICE HEADER(NSH), VNF SUPPORT MPLS



Push label stack for Service Chaining.

VNF support MPLS label

TELCO CLOUD

WHAT IS THE TELCO CLOUD ARCHITECTURE? HIGH LEVEL ARCHITECTURE

5

Key Properties

1. Physical distribution providing fungible cloud resources close to Telco consumer and business eyeballs.
2. Enables applications to have:
 1. Low Latency
 2. High Availability (through distribution)
 3. High volume of last mile throughput; minimizing network wide capacity growth (choke points)
3. Seamless Integration of DC and WAN technologies leveraging existing network and operational procedures.

Connectivity Building Blocks

WAN/METRO

BGP (Control Plane)
MPLS (Service)
MPLS (Transport)

+

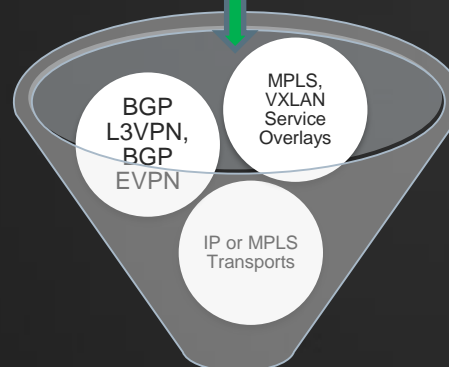
DC Fabric

BGP / OSPF (Control Plane)
IP (Transport)

+

DC Overlays

MPLS, VXLAN, IP, GRE, etc.



Openstack

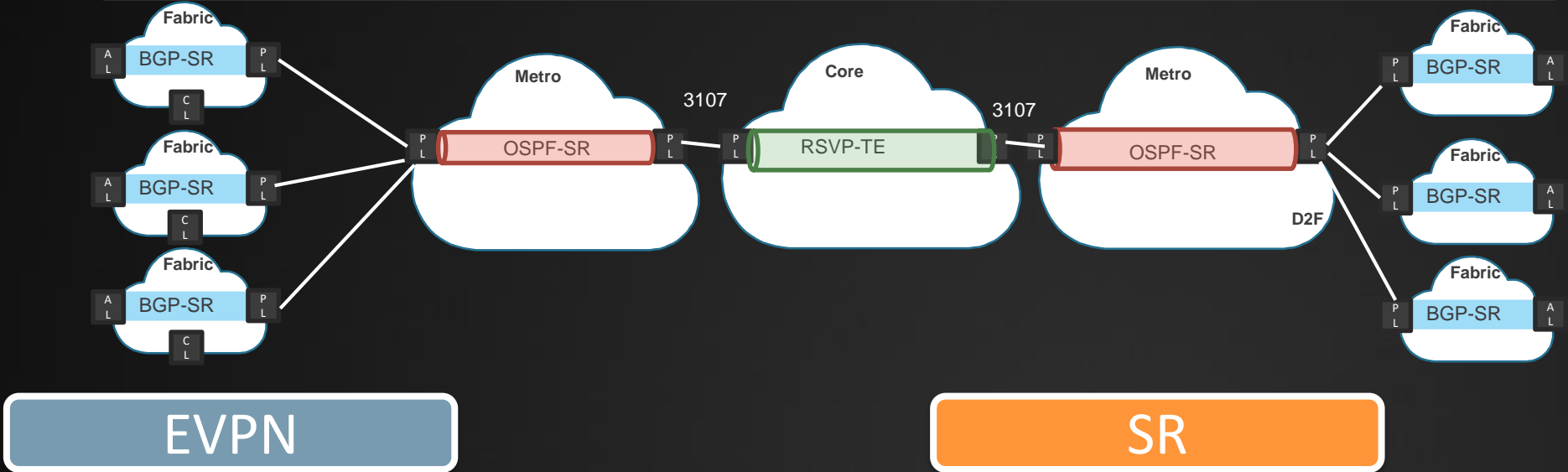
Neutron & Neutron Extensions,
etc.

Telco Cloud



TELCO CLOUD HIGH LEVEL REQUIREMENTS 10K FEET

MPLS in SP Fabrics - High level vision

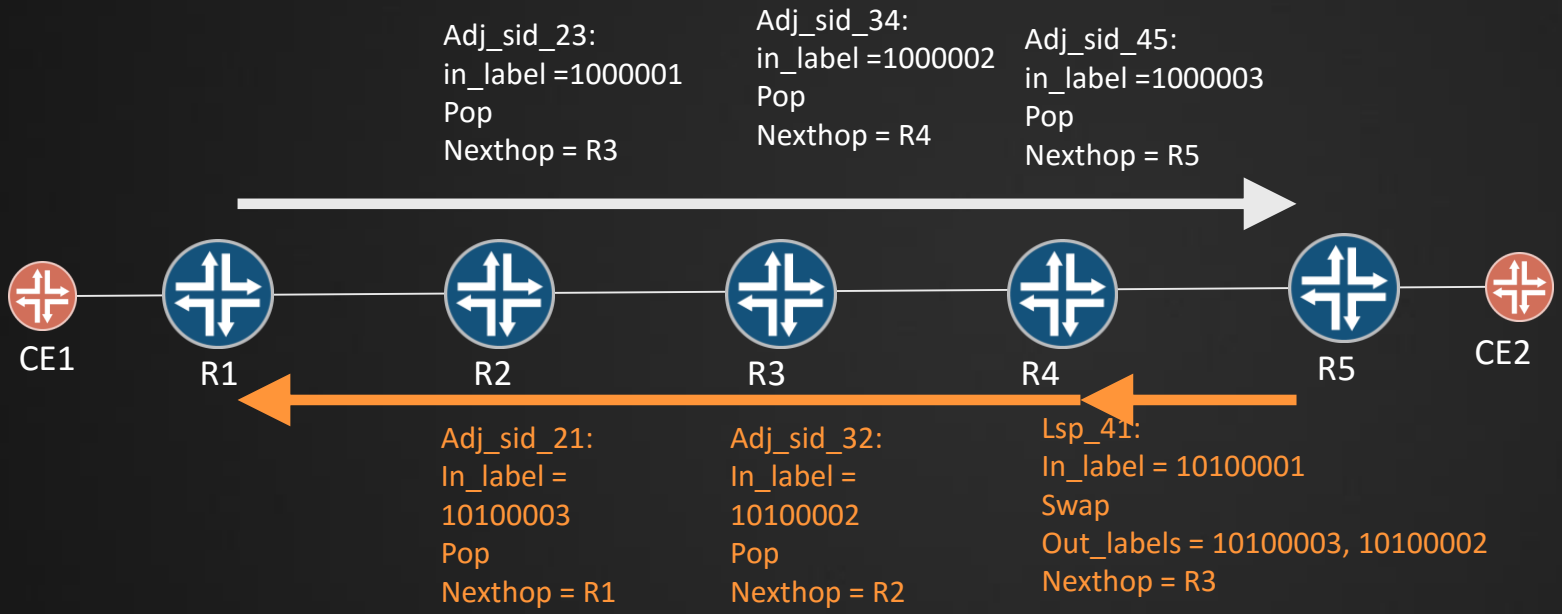


- EVPN Signaling is a key requirement for **all** control plane signaling
 - EVPN-VPWS with flexible-cross-connect for all L2 pseudowires
 - EVPN-MPLS multi-point with IRB
 - EVPN-VXLAN for IP fabrics

- Underlay transport is based on Segment Routing
 - No IGP in Telco Cloud. Only BGP-LU with prefix-SID extensions
 - Metro moves to OSPF-SR

STATIC SEGMENT ROUTING

Step1: Build the Segment Routing Topology, Single Hop LSP



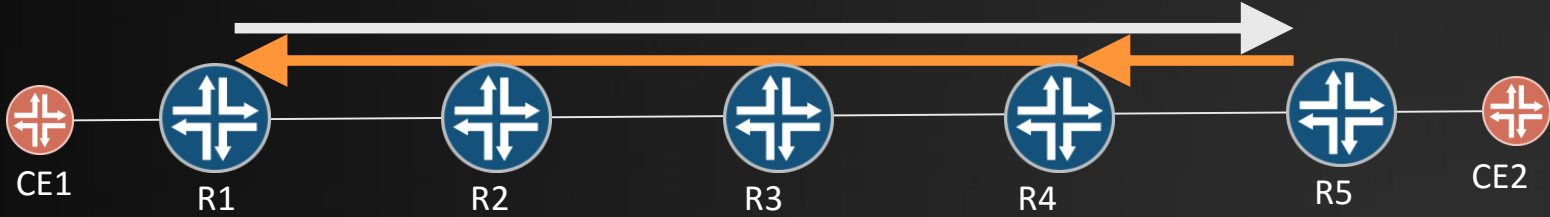
STATIC SEGMENT ROUTING

Step2: Push the SR LSP from Edge



Lsp_15:
Dest = R5
Push
Out_labels = 10000003, 10000002, 10000001
Nexthop = R2

Lsp_51:
Dest = R1
Push
Out_label = 10100001
Nexthop = R4



Adj_sid_23:
in_label = 1000001
Pop
Nexthop = R3

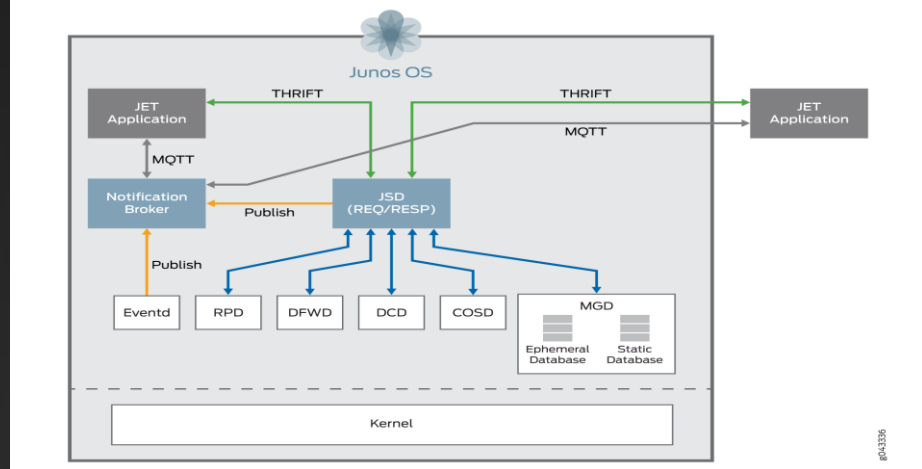
Adj_sid_21:
In_label = 10100003
Pop
Nexthop = R1

Adj_sid_34:
in_label = 1000002
Pop
Nexthop = R4

Adj_sid_32:
In_label = 10100002
Pop
Nexthop = R2

Adj_sid_45:
in_label = 1000003
Pop
Nexthop = R5

Lsp_41:
In_label = 10100001
Swap
Out_labels = 10100003, 10100002
Nexthop = R3



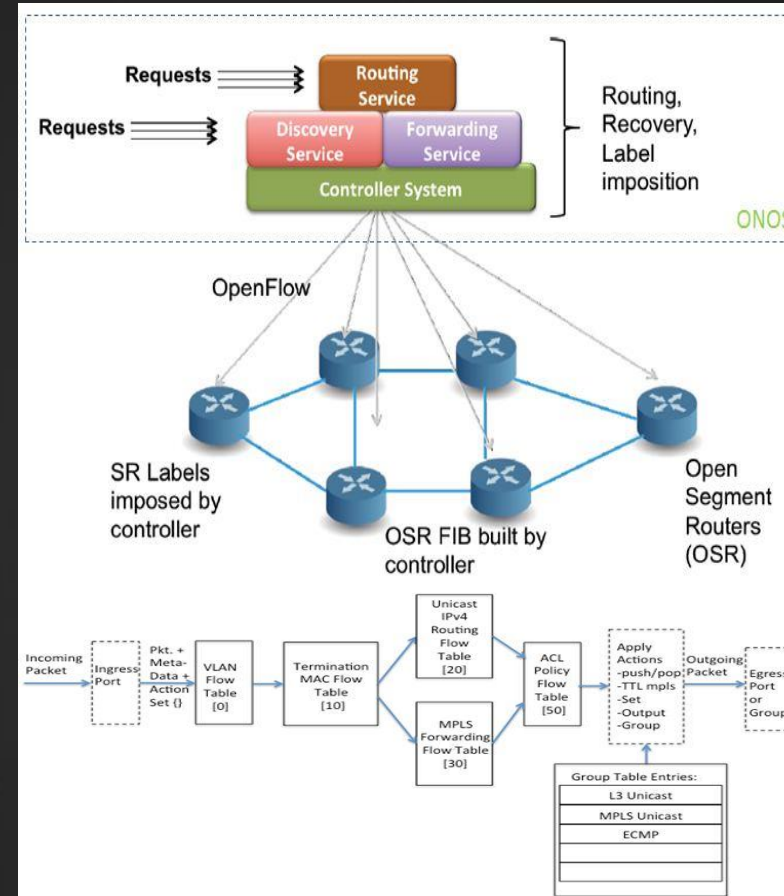
```
Ingress LSP with a stack of Adj-SID labels:
destJnxP = IpAddressAddrFormat("128.9.148.133")
dest = JnxBaseIpAddress(destJnxP)
lsp = RoutingStaticLspEntry()
lsp.name = "lsp_15"
lsp.type = 0 << ingress
lsp.Prefix = StaticLspEntryPrefix()
lsp.Prefix.destination = dest
lsp.label_operation = 0 << push
lsp.outgoing_labels = ["1000003", "1000002", "1000001"]
lsp.nexthop = "55.1.12.2"
lsp.preference = "6"
lsp.metric = "1"
addReq = RoutingStaticLspAddRequest(lsp)
addReply = staticLsp.StaticLspAdd(addReq)
print 'Reply status = ', addReply.status
```

OPENFLOW WITH SEGMENT ROUTING

ONF's SPRING-OPEN



- OpenFlow 1.3.4 can push 2 labels
 - Service label and Tunnel labels
 - Use Openflow group Chain to push multiple labels
- Openflow Build the Segment Routing Topo
 - Adj SID for POP
 - Node SID for continue(no change/no swap)
- No RSVP-TE/LDP and IGP on those routers
 - Only MPLS dataplane and Static configure from Openflow
- A lot of limitations BUT can show
 - Intelligence on Controller, very ugly CLI on Controller
 - White Label box with simple MPLS forwarding Plane
 - Demo in Dec 2014. <https://goo.gl/ddeX5N>



AGENDA

Introduction

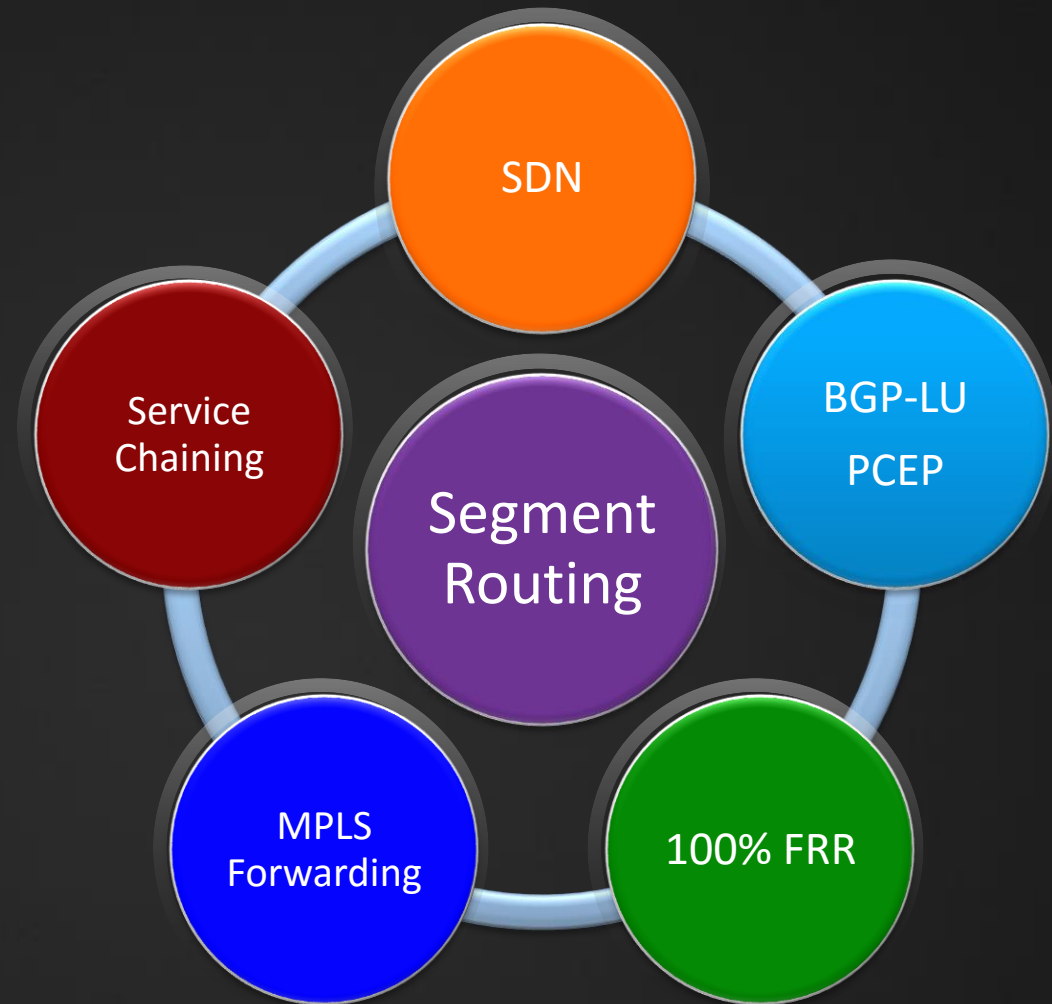
Segment Routing Deep Dive

Segment Routing SDN and Use Case

Summary

Summary- Segment Routing Re-Invent MPLS

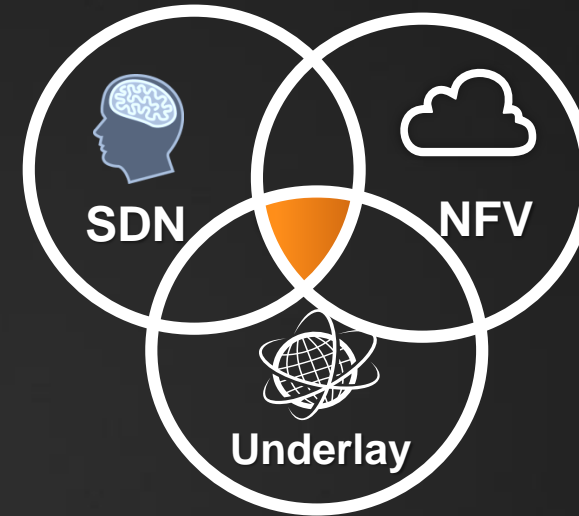
- Seamless work with SDN, BGP-LU/PCE-P Architecture. instantly tunnel setup. for next generation Application driven networks
- Work with NFV, such as Service Chaining
- Simplified MPLS Control Plane, OSPF/ISIS only. No need Signaling for tunnel setup. Tunnel path decided by ingress router.
 - source routing and hence explicit routing
- less status inside network element(router/switch)Topology based on Adj/Nodal information. Independent with Application Status
- 100% IP fast reroute protection, Fit for any topology
- Work great with Traffic Engineer and IPv6.. With QoS, OAM/SLA



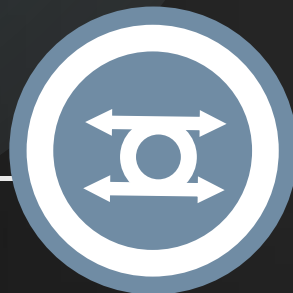
Segment Routing Customers

Re-invent MPLS again! Foundation of NFV/SDN

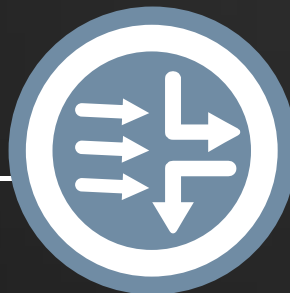
- Major vendors claim to support, ALU/Cisco/Huawei/Juniper
- Known customer transforming to SPRING
 - AT&T CORD
 - Microsoft SWAN
 - China OTT, Tencent/Alibaba
 - Japan Softbank/NTT
 - ANZ Telstra etc



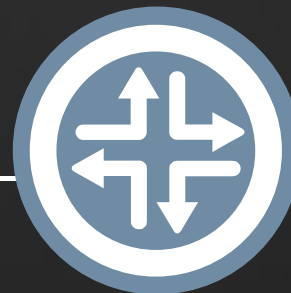
CPE



Access



Edge

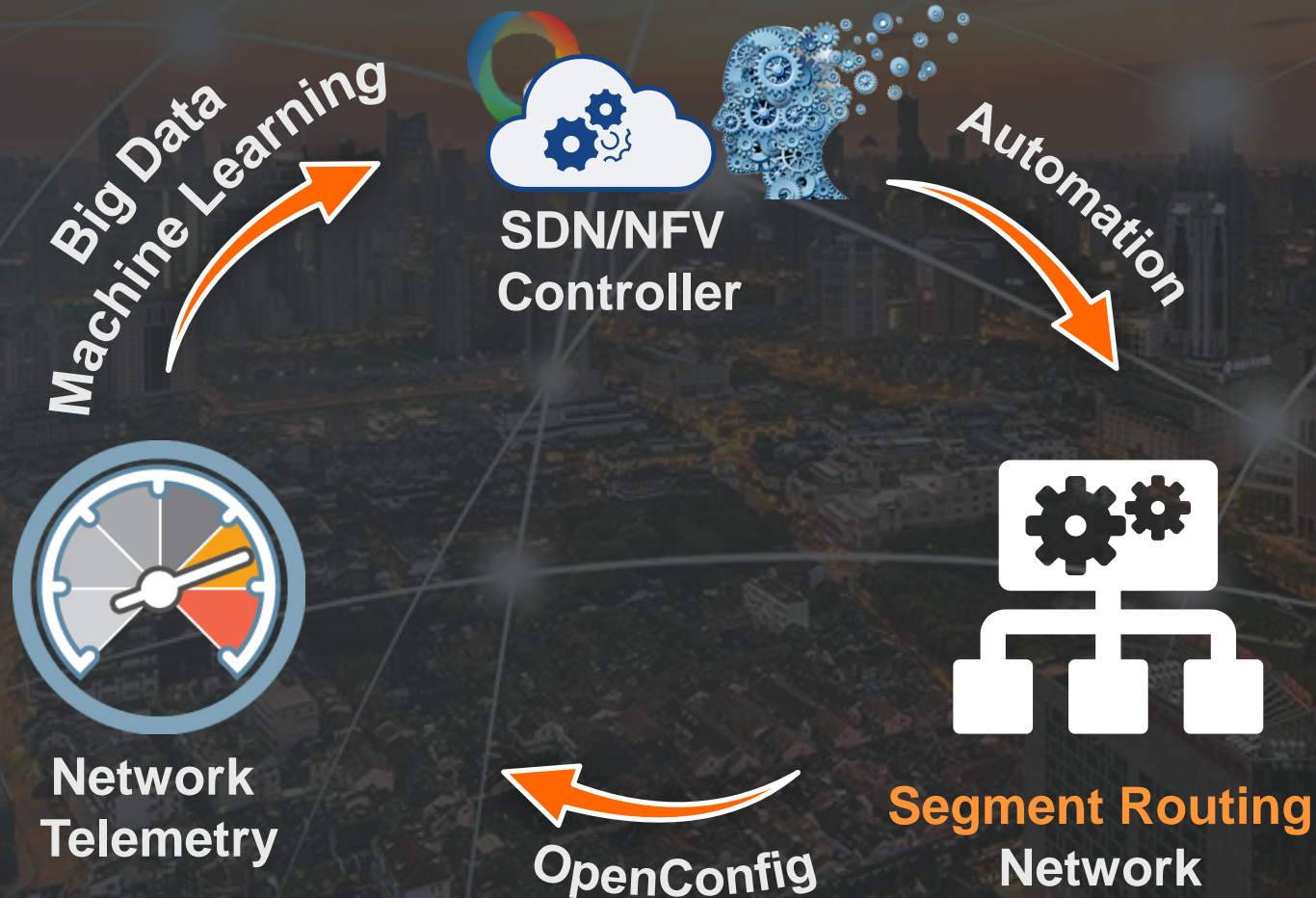


Core



DC

ROAD TO SELF DRIVEN NETWORK

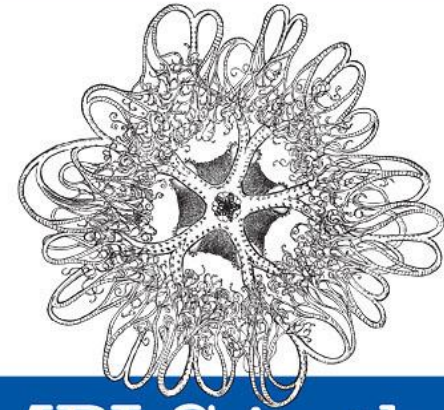


SUMMARY

- 1 Segment Routing Design for SDN
- 2 Segment Routing simplify Protocols
- 3 Segment Routing enable better traffic engineer, IGP/BGP, Egress Peering Engineering
- 4 Segment Routing Provide better FRR protection
- 5 Segment Routing can be deployed in All Domains, DC, Metro, Access, Telco Cloud etc.

THANK YOU

O'REILLY



MPLS in the SDN Era

INTEROPERABLE SCENARIOS TO MAKE NETWORKS
SCALE TO NEW SERVICES

JUNIPER
NETWORKS

Antonio Sánchez-Monge &
Krzysztof Grzegorz Szarkowicz